DOCUMENT RESUME

ED 178 299 SE 028 842

AUTHOR Haan, Norma
TITLE An Exploratory Investigation of the Effects of an Initial Experience with SCIS's Material Objects Unit on First Grade Children and Their Teachers.
INSTITUTION Far West Lab. for Educational Research and Development, Berkeley, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Mar 68
NOTE 75p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Attitudes; Cognitive Development; *Educational Research; Elementary Education; *Elementary School Science; Questioning Techniques; Science Course Improvement Project; *Science Curriculum; *Science Instruction; Scientific Concepts; *Teacher Behavior
IDENTIFIERS *Science Curriculum Improvement Study

ABSTRACT

This report presents the results of an investigation of the Science Curriculum Improvement Study (SCIS) program and the effects of the Materials Objects unit on first-grade students and teachers. The contents of the report include purposes and limitations of the study, discussion of the instruments, observers, reliability of the Q sort, the differences between the experimental and control classrooms, the SCIS tasks, the attitudinal measures, and recommendations based on the results. The limitations of the study are also noted. (Author/SA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178299

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

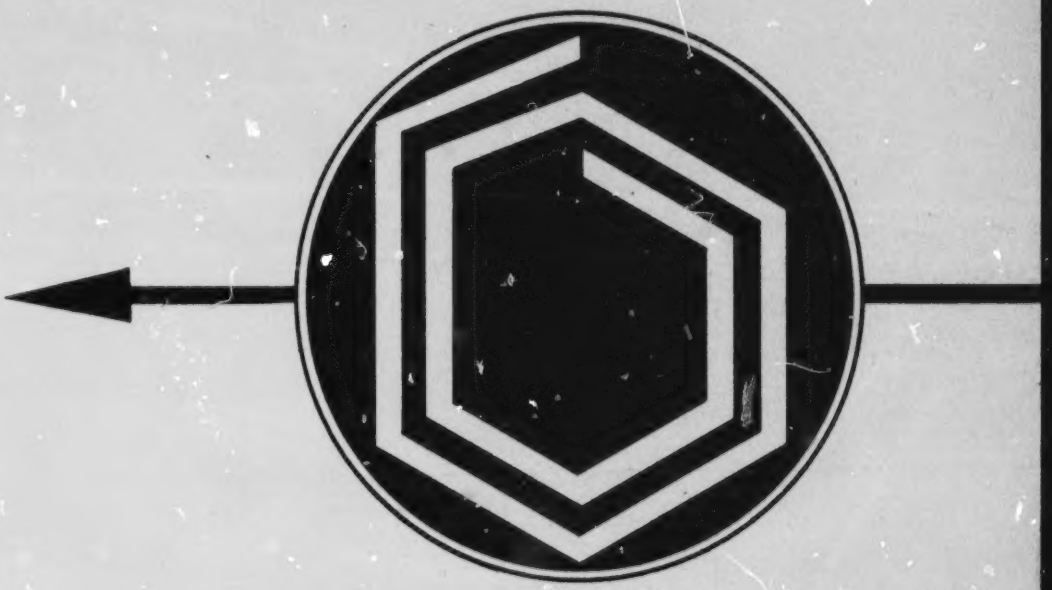
Far West Lab.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: SE PS

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view. SP



AN EXPLORATORY INVESTIGATION OF THE EFFECTS OF AN INITIAL EXPERIENCE WITH SCIS'S MATERIAL OBJECTS UNIT ON FIRST GRADE CHILDREN & THEIR TEACHERS

By NORMA HAAN

PUBLISHED BY
FAR WEST LABORATORY FOR EDUCATIONAL RESEARCH AND DEVELOPMENT
1 Garden Circle, Hotel Claremont, Berkeley, California 94705

SE 028 842

THE FAR WEST LABORATORY FOR EDUCATIONAL RESEARCH AND DEVELOPMENT is a regional educational laboratory, established through a Joint Powers Agreement in February, 1966. Present signatories include the Regents of the University of California, the California State Board of Education, the Trustees of the California State Colleges, the County Superintendent of Schools of the County of Monterey, the Board of Education of the San Francisco Unified School District, the Regents of the University of Nevada and the Nevada State Board of Education.

Published March 1968. Work reported herein was performed pursuant to a contract with the U. S. Department of Health, Education and Welfare, Office of Education.
This publication should not be reproduced without prior consultation with the Far West Laboratory for Educational Research and Development, 1 Garden Circle, Hotel Claremont, Berkeley, California 94705 (Telephone 415-841-9710).

TABLE OF CONTENTS

I. INTRODUCTION

II. PURPOSES OF THE STUDY

III. LIMITATION OF THIS STUDY

IV. THE CLASS OBSERVATION

- A. The Instrument
- B. The Observers, Their Training, and The Scheduling of Observations
- C. The Reliability of the Q Sort Taken as a Whole
- D. The Criterion Sort by the SCIS Staff

V. COMPARISONS BETWEEN THE CLASSROOM SORTS

- A. The Differences between the Experimental and Control Classrooms
- B. The Changes within the Experimental and Control Classrooms
- C. The Effects of Experience

VI. THE CHILDREN

- A. The SCIS Tasks
- B. The General Cognitive Tasks
- C. Locus of Control
- D. Two Other Attitudinal Measures
- E. Children's Questions with Regard to the Curious Objects

VII. DISCUSSION OF RESULTS

VIII. RECOMMENDATIONS

AN ABSTRACT

An Exploratory Investigation of the Effects of an Initial Experience with the Science Curriculum Improvement Study's Material Objects Unit on First-Grade Children and Their Teachers

In December, 1966, an agreement was reached among the Science Curriculum Improvement Study (SCIS), the Berkeley Unified School District (BUSD) and the Far West Laboratory for Educational Research and Development, that the Laboratory would conduct an exploratory investigation of the effects of the new Material Objects unit upon first-grade students and teachers in one Berkeley school and that a comparable group in another school with a similar child population would participate as a control school.

The project began in January, 1967, with the SCIS staff conducting training sessions for the teachers of the experimental school and the Laboratory pretesting the children in the first-grade classes of both schools. SCIS staff continued in a variety of ways to work with the experimental teachers throughout the spring semester.

A group of experienced teachers, most of whom were doing graduate work in education at the University of California or San Francisco State College, were employed by the Laboratory to observe both groups of teachers several times during science instruction periods - the experimental group teaching the Material Objects unit and the control group teaching science according to their usual procedures. Their observations were recorded by a *Q* sort which was specially designed for this study to describe instructional behaviors that are relevant to the SCIS curriculum. Two other kinds of teacher observational measures were obtained to provide a frame for understanding the experimental teacher's use of SCIS instructional behaviors: (1) a group of ten teachers from three non-study schools, each with one to four years of experience with the program, were also observed; (2) twelve of the SCIS staff used the *Q* sort to describe a theoretical teacher, the SCIS criterion teacher who would maximize the specific intentions of the program in her instructional behaviors.

In May the data-collection phase of the project was completed by a second testing of the children in both schools. Assessments of the children's gains were made in several ways: their improvement in understanding the concepts embedded in the Material Objects unit; their general cognitive growth; their attitudinal shifts with regard to themselves and adults; and their degree of curiosity. The children's data were analyzed by a comparison of changes.

Various analyses of classroom observations were made. The intention was to understand tentatively specific teacher effects upon children's utilization of the experiences arising either from experience with SCIS or from initial conformance of teaching processes to the program's

expectancies. Some attention was also given to shifts in the experimental teachers' behaviors as they became more familiar with the program during the semester.

In brief, the results show that children in the experimental school made some attitudinal shifts toward conceiving of themselves as able to determine their own fate. But greater effects upon children occurred when their teachers employed teaching strategies which were in accord with the SCIS's staff criterion descriptions. These results include improved performance in an SCIS-derived task and in cognitive tasks of conservation and understanding gravity, a greater sense of determining one's own fate, and a change in questioning activity toward more complexity, risk-taking, and imaginativeness. Some teachers showed shifts related to the SCIS experience but others did not. The several limitations of the present study are noted and discussed.

INTRODUCTION

A preliminary commercial version of the Material Objects unit for use with first grade children was released by the Science Curriculum Improvement Study (SCIS) in June, 1966.[1] Different units in physical and biological sciences and others for older children are in various stages of development. From the time of a National Science Foundation grant in 1959, the staff of SCIS, under the direction of Dr. Robert Karplus, has worked to develop a sequential, articulated science program for elementary schools. During this period a number of schools and teachers have worked with SCIS in trying out various units and providing corrective and useful feedback to the developers.

The SCIS program explicitly recommends that teachers reinterpret their traditional classroom role so that they function more as guides, facilitators, and attentive observers of children's learning than as instructional authorities. The curriculum guide makes it clear, however, that teachers should not take a passive position when children make errors of fact, and this recommendation, among others, differentiates the program from earlier "activity" programs, which also emphasized child involvement. The SCIS definition of teacher role logically follows from its conceptualizations of the particular aspects of children's learning that seem to be paramount and critical. In this regard SCIS not only has been attentive to the large and overgrowing body of research in cognitive development, incited by the earlier work of the Swiss psychologist, Piaget, but also has been influenced by a general model of scientific investigation. The structure of the Material Objects unit reflects the conviction that the young child first thinks concretely, and only later moves to meaningful abstractions by accumulating and understanding a series of concrete experiences whose meanings and implications are made inescapable to him by his own observations. This position draws a sharp distinction between the child's producing a "correct" answer upon request and being convinced of the truth of his solution. One interesting result which is logically congruent with this

---

[1]This investigation was conducted and largely financed by the Far West Laboratory for Educational Research and Development according to an agreement with the Science Curriculum Improvement Study and the Berkeley Unified School District. The Science Curriculum Improvement Study took financial responsibility for the observation of experienced SCIS teachers in three additional schools. The author is indebted for the helpful criticism of an earlier version of this paper by John Hemphill, Paul Hood, Robert Karplus, Ellen Siegelman, and Herbert Thier.

expectancy was reported by Siegelman (1957). In a recent SCIS-conducted evaluation, children with SCIS curricular experience responded more frequently with "I don't know" when they probably did not know; whereas children without SCIS experience tended to make improbable guesses. Children with SCIS experience were more likely to present plausible hypotheses to puzzling situations. Thus, SCIS would have the teacher take the role of stimulating the child's curiosity so that he exposes himself, takes the risk of questioning, and thereby draws conclusions in his own terms.

Since the SCIS curriculum intends that both teachers and children assume non-traditional roles, and since its full effectiveness may depend upon this shift, an investigation of the outcomes should attempt to measure the less tangible, but nevertheless important changes in both teachers and children--in addition to the more obvious and expected gains in the children's understanding of the scientific concepts directly presented in the unit. If teacher and children do not adopt these suggested roles, it would be clear that the group had not actually experienced the curriculum as it was intended, so that the results in terms of child effects would have to be understood in a different light. As early as 1939, Jersild and his colleagues (Jersild, Thorndike, Goldman, & Loftus, 1939) found that some teachers did not make behavioral shifts recommended by a new "activity" program. However, it appears that the SCIS curriculum makes a considerable advance over much earlier programs, which were more often exhortatory than specific. It presents an abundance of specific materials and procedures which may provide the vehicle for all classroom participants to move toward a situation where children are more actively involved in their own learning. It is reasonable to expect that the SCIS curriculum will have less dramatic, but possibly more durable and generalizable effects upon children than some new curricula which intend to have direct and dramatic impacts, since it takes the position that indirect influence upon children is more likely to promote certain learning and curiosity.

PURPOSES OF THIS STUDY

The present investigation was primarily designed to study the effects of one semestr's experience with the Material Objects unit on the first grade children and teachers in one school compared with a similar group in another school where the usual science instruction procedures were followed.

Since both of these schools serve populations which are relatively economically disadvantaged, the present study affords some opportunity to understand the effects of this curriculum within this context. Some additional study was made of another group of teachers who had a longer experience with the curriculum. More specifically, the investigation was directed at two sets of questions: (1) Do the first grade children in the experimental school, when compared to those in the control school, show greater gains in assimilation of the scientific concepts presented in the Material Objects unit? Do they show greater cognitive gains generally? Do they show greater attitudinal shifts toward themselves and their teachers? Can the differences in the amount and kind of gains among the experimental children be attributed to differences among their teachers' adaptation of the SCIS role? (2) Do the teachers in the experimental school change from

their baseline performance following experience with the training and, perhaps more importantly, the teaching of the curriculum itself? If there is a change, does it represent a convergence with the curriculum's expectancies? In what way do the procedures of the experimental school teachers differ from those of the control school teachers who had no SCIS experience? Do the teachers in the experimental school, who had only one semester of experience with SCIS, differ in their procedures from teachers who have had a longer experience? The teachers in the experimental school participated in 16 hours of training in group sessions and a SCIS supervisor visited their rooms approximately every other week.

LIMITATIONS OF THIS STUDY

Before the discussion turns to methods used and results obtained, several defects in this study must be stated which not only limit the applicability and generalizability of its results but in some cases also decreased opportunities for uncovering effects which may very well have existed. (1) The first is a commonplace one in school research: most statistics assume that samples are randomly drawn; the present one was not; but it was not possible or even reasonable to select such samples. (2) Group testing was used, even though it is clear that more solid measures can be obtained from six- and seven-year-old children by individual testing. The decision for group testing was dictated by the limitations in funds, the brevity of the period available for planning the study, and the exploratory nature of the investigation. If the present measures have less than optimal reliability because of child inattention, this deficiency may be compensated for by the comparatively larger number of children in the sample (a total of 167 were both pretested and posttested). (3) No reasonable control of the children's ability level could be obtained. (Some of the children in the experimental group previously had had a group I.Q. test, but none was available for the control children; a number of different reading systems were being used in the two schools, so that reading level could not be used as an alternative index). (4) Of the achievement-type tests used in the pretesting, two proved to have too low a ceiling to permit the high-performing children to gain. Two others are apparently so generally sensitive to growth processes in six- to seven-year-old children that substantial portions of the subjects attained perfect or near-perfect scores on the posttests. This occurrence makes it almost impossible to uncover consistent gains when they do exist. The reader will recognize that there is a noticeable paucity in the means of measuring many of the positive aspects of children's growth as soon as one departs from conventional I.Q. and achievement tests. Clearly SCIS and a number of the other new curricula intend to accomplish different goals with children. Most often these emphasize "thinking" instead of fact accumulation. Hence, new and broader intentions necessitate different kinds of assessments; their development will inevitably entail a number of clumsy successes and outright failures.

Another inelegancy needs to be noted: some additional observations were made of ten experienced SCIS teachers. These were not randomly drawn from all available SCIS teachers as would have been desirable. Instead they were chosen because their schools were in convenient geographic locations, and their teaching procedures were well known to the SCIS staff. Furthemore,

most of them were not first grade teachers, but were second- and third-grade teachers. They are included in some of the present analyses, however, because of the critical need to obtain more understanding of teacher reaction to the experience and of teacher change in general.

One last limitation must be discussed with regard to the timing of the new experience and its observation. The study began in the middle of the year so that most children would already have formed their attitude toward learning within their own particular classrooms. Most developmental studies have found that children are most fluid and changeable at the times of radical shifts in their environment (cf. Block and Haan, in press, for a discussion of this point). Thus, larger change effects might have been discovered if the study had begun in the Fall. Moreover, the late beginning meant that none of the teachers had completed the activities of the Material Objects unit at the time of the last observation; most of them were only about two-thirds finished, and they seemed to be experiencing some personal pressure to finish.

The SCIS curriculum has in general two different kinds of activities--called invention and discovery lessons. Discovery lessons involve children primarily in individual activity and teachers usually take an assisting, supportive role with individuals. During invention lessons, teachers generally take more active, stimulating, leading roles. Of course, different teachers interpret these two kinds of lessons in individualized ways. An optimal study design would have made provision that all the teachers be observed when they were involved with the same kind of lesson (or some other controlled variation). The pressure that such a plan would have exerted on both teachers and observers made it seem wise to abandon this aspiration for methodological purity. The discussion at the end of this paper will assess the results within the context of these various constrainsts.

THE CLASSROOM OBSERVATION

The Instrument

A Q-sort deck composed of 66 items was specifically designed for this study which included teacher behavior and child reactions important to the SCIS goals.[2] A listing of the items if found in Appendix A. Their reliabili-

---

[2]The final 66 item deck had a number of authors: an original pool was constructed by Norma Haan and Aubrey Haan along with several San Francisco State College faculty members; later Norma Haan and Peter Lenrow revised the items for a nursery-school setting; then Ellen Siegelman of SCIS and Norma Haan revised the deck again with other SCIS staff members' advice; finally the observers who participated in this study made suggestions for refining and clarifying the present items, and these were also incorported.

ties, after the Spearman-Brown correction was calculated for the first two obe.vation periods, is cited in Appendix B.[3] This *Q* sort should be regarded as a research instrument which is subject to a number of vagaries that will be noted below. A cluster analysis by the Tryon method was made of individual observers' sorts; this is not a proper procedure because each individual sort was not independent of all others, observers having worked in pairs and the same teachers having been seen successively. The small number of truly independent observations made proper clustering or factoring impossible. The present clustering was done, with due recognition of its impropriety, to obtain a rough and tentative idea of the major, empirical dimensions in the *Q* sort. The results can be found in Appendix C.

This method of attaining classroom descriptions was chosen for a number of reasons. Since the SCIS program specifically expects teachers to assume a less directive and verbalizing role, and many existing measures are based upon teacher verbalization, it was likely that teachers who took the SCIS role would be unreliably observed by extant methods. Moreover, the degree of teacher verbalization with first graders is undoubtedly less than in upper grades. Global evaluative ratings did not seem appropriate for the purpose, since there was no way that the intelligent observers could be kept in a naive state with regard to the purposes of the study. Some degree of impartiality was probably insured, however, since the observers were working for the Laboratory and were unacquainted with the aims of SCIS. They were told during training that the *Q* sort was to be used to describe the events within classrooms and that an overall judgment of the teacher was not of interest.

There were, moreover, a number of positive reasons for using the *Q* sort in this study. Its exploratory nature and SCIS's expectancies that a wide variety of both tangible and less tangible teacher-child interactions characterize the optimal situation mean that an assessment model of the real situation should attempt to represent an equal complexity. Research in teacher effectiveness seems to have taken one of two approaches when it has been confronted with the complexity of classrooms: the unit of assessment is often made narrower and more concrete, so that the observer simply records a number of predetermined behaviors by tally marks, or sometimes, a small number of ratings are made on a global basis according to the personal evaluations of observers. The judgments of such observers are clearly value

---

[3]In brief, this *Q* sort involves a fairly large sample of items, in this case 66, which are sorted into nine numbered categories along a continuum from "most descriptive" to "least descriptive" of the observed classroom situation. The distribution is forced into a symetrical semi-normal one by restricting the number of items which may be placed in each category. In this case the distribution frequencies were 5, 6, 8, 9, 10, 9, 8, 6, 5. Individual items are scored according to the category number (1-9) of their placement and the distribution of 66 scored items, the *Q* sort, is correlated with other *Q* sorts to determine the extent of agreement. These values are expressed as Pearson product-moment coefficients.

matters (e.g., what is an orderly or disorderly classroom). Judgments are usually "calibrated" in a series of training sessions where junior observers are trained to agree with a senior investigator's evaluations. The Q-sort technique, however, permits its user to describe a situation (traditionally it has been a person) globally, by placing a large number of relatively precise items within a differentiated and interrelated framework. Moreover, the sorter's job is limited to positioning variables within the framework of describing a single, observed classroom, and this aspect of the task relieves him of the necessity of comparing a particular classroom with the theoretical "all other" classrooms. These features of the Q sort made it seem particularly useful for this investigation: complex behavior needed to be assessed and there was little reason to think that a predetermined and narrow selection of dimensions would necessarily and relevantly describe the particular kind of teacher behavior and classroom interaction that was related to the desired criterion of child gains.

The Observers, Their Training, and the Scheduling of Observations

Eight observers, all of whom had had teaching experience, mostly in the primary grades, were trained in a group. They watched a first grade teacher in a non-project school teach one lesson of the Material Objects unit (about 40 minutes), and then they completed a Q sort. Discussion of discrepancies in item placements then took place among the group which included the investigator and a SCIS staff member. This procedure was repeated once again with the observation of another teacher. The investigator took the position during group discussion that bona fide disagreement could be expected, since one observer might catch something that another missed and be more impressed by some behavior of these directive SCIS teachers than another was. Group discussion was directed toward achieving a consensus with regard to the meaning of the words in the items. In other words, a deliberate attempt was made to avoid training the observers to achieve a "value" agreement. Instead, the intent was to obtain a semantic consensus with regard to the meaning of items. Experienced teachers were chosen because of the supposition that the perceptions of less seasoned observers are undifferentiated; they often fail to make the connection between a teacher's purposes and her behavior.

The schedule for the observations of classrooms in the control and experimental schools was so arranged that no observer ever saw the same situation twice. Originally, it had been planned that three observations--beginning, middle and terminal--would be made of each classroom by a pair of observers, but due to the pressure of school closing for the summer recess and unanticipated events in the control school, the third observation was not made there. Furthemore, an attempt was made to vary the pairing of observers, so that whatever idiosyncratic biases a pair of observers might share would be randomized rather than compiled. The personal exigencies of both teachers and observers made it impossible to achieve this goal completely, but out of the 35 pairings of observers, repetition occurred only 11 times. Although the non-repetitive scheduling and the variation in observer pairing throw some idiosyncratic error into the measures obtained, this procedure has the more crucial advantage of providing greater

certainty that findings are veridical instead of an accumulation of bias. Classrooms in both schools were observed only during science instruction periods: the experimental children with SCIS and the control children with curricular material chosen by their teachers. (Two basic series were available in the control school: D. C. Heath's Basic Science Series by Schneider; and Allyn and Bacon's Exploring Science by Thurber.)

The Reliability of the *Q* sort Taken as a Whole

The congruence between pairs of observers in their total description is another kind of reliability that is important to the *Q* sort. Some studies involving classroom observation are also concerned with the congruence between one time and another; this is not an expectancy in our study since the new experiences of the teachers and children in the experimental school were specifically intended to produce change in their interaction. Table 1 shows some evidence for change in the experimental classrooms and consistency in the control classrooms. The latter were more consistent between the first and second observations than were the experimental classrooms--a finding that is even clearer when one teacher (H), who was not reliably observed at either the first or second period, is discounted.

The raw agreements between two observers and the reliability of their *Q* sorts after compositing is also shown in Table 1. Examination of this information indicated that the following interpretations can be entertained: overall reliability is sufficiently satisfactory to warrant making composite *Q* descriptions of the six independent observations for the three situations (or four for two situations in the case of the control school) in order to achieve a single set of scores to relate to the SCIS criterion sort which is described below. It further appears that a longer training experience for the observers would have been beneficial since their level of reliability improved, at least moderately, between the first and second observations (this possibility agrees with observers' comments to the effect that they held the items better in mind by the time of the second observation).

The same calculations as above for the experienced teachers are shown in Table 2. It is of some interest that the observers' perceptions of these classrooms were generally more congruent and that the two situations which did not produce substantial agreement, P and R, were not high criterion SCIS classrooms, either by the SCIS staff rating or by correlation with the criterion sort. These two teachers also had less experience. Evidently the curriculum and their teaching styles are not particularly compatible. In general there is good agreement between the staff rating and the criterion sort. This generally high level of agreement fulfills an obvious expectancy that more experience with the SCIS permits a teacher to behave more certainly and clearly.

The Criterion Sort by the SCIS Staff

Twelve SCIS staff members used the same *Q*-sort deck to describe the hypothetical teacher and classroom interaction which they thought would optimize children's learning within the SCIS experience. Their agreement

Table 1

Observers' Agreements, Reliabilities, and Relationships to Criterion Sorts[1]

(Experimental and Control Schools)

	Observers' Raw Agreements			Reliability of Composite				Agreement between Composites			Agreement between Composite and SCIS Criterion			
	Observation period			Observation period			Overall	Observation period			Observation period			
	1st	2nd	3rd	1st	2nd	3rd	Six Observers	1st-2nd	1st-3rd	2nd-3rd	1st	2nd	3rd	Overall
Experimental														
A	.15	.56	.60	.27	.72	.75	.53	-.04	-.14	.42	.30	.05	-.38	-.04
B	.48	.47	.26	.65	.64	.41	.60	-.04	.38	.26	.40	.29	.62	.64
C	.49	.67	.49	.66	.80	.66	.63	.66	-.14	-.08	.74	.74	-.18[3]	.67
D	.61	.75	.58	.76	.86	.74	.91	.78	.65	.76	.59	.74	.52	.68
E	.32	.40	.60	.48	.57	.75	.74	.17	.59	.40	-.53	-.02	-.46	-.43
F	.53	.56	.46	.69	.72	.63	.81	.55	.39	.60	.02	.25	.43	.28
Control														
G	.55	[2]		.71	[2]		.76				-.38	[2]		-.02
H	.13	.15		.23	.27		.37	.20			-.13	.28		.10
I	.51	.65		.68	.79		.85	.71			.64	.80		.78
J	.56	.32		.72	.48		.79	.68			-.64	.47		-.60
Overall $\bar{X}$[4]	.45	.52		.61	.68		.73	.46						
Experimental $\bar{X}$	.44	.58	.50	.60	.73	.67	.73	.40	.32	.43	.29	.39	.11	.35
Control $\bar{X}$	.45	.40		.61	.55		.73	.57			-.13	.28		.10

(1) Control school did not participate in third observation; (2) one observer missed appointed time; (3) large proportion of new children in room; (4) all means are calculated by the r to z transformation.

Table 2

Observers' Agreements, Reliabilities, Relationships to Criterion Sorts and SCIS Staff Ranks:

Experienced Teachers

	Observers' Raw Agreements	Reliability of Composite	Correlations Observers' Composite and SCIS Criterion	SCIS Staff $\bar{X}$ Rating	SCIS Staff Range
L	.65	.79	.80	4.5	4-5
M	.73	.84	.79	4.7	4-5
N	.64	.78	.74	4.0	3-5
O	.77	.87	.83	4.0	3-5
P	.22	.36	.11	3.3	3-4
Q	.66	.80	.80	5.0	5
R	.20	.33	.37	3.7	3-4
S(2)	.46	.63	-.18(2)	4.5	4-5
T	.75	.86	-.79	2.2	1-3
U	.55	.71	.64	4.8	4-5
Overall $\bar{X}$ =	.59(1)	.74	.51		

(1) Calculated by r to z transformation.

(2) Both observers made a note that teacher did not take an active role during observation.

among each other was high, ranging from .64 to .89 with a mean of .78, calculated by the *r to z* transformation. All of these individual sorts were composited to achieve a single criterion sort; the reliability of this composite was .98.

Readers will, of course, be interested in the content and scores for the SCIS criterion sort. Table 3 presents the items according to their characteristic and uncharacteristic position; these serve to define the curriculum as it is envisioned by the SCIS staff. Note that the most characteristic items show the staff's strong focus upon children's cognitive growth, its commitment to a guiding rather than controlling role on the part of the teacher, and the intention that children must take responsibility for their own learning.

Perhaps the most interesting information in Tables 1 and 2 is that several inexperienced teachers, C, D, and I, were relatively close to the SCIS criterion before training started (of course control teacher I did not participate in any training) and experienced teachers, P, S, and T, despite the opportunities afforded by the structure and materials of this curriculum, interacted with children in a manner that was incompatible with its aims. Two experimental school teachers, B and F, appear to have moved toward the ideal description, but teachers A and E seem not to have been much affected.

A high score on the criterion sort implies some degree of homogeneity in the ways that effective SCIS teachers proceed, but a low score can clearly be achieved in a variety of ways. Some attempt is made here to discover the attributes of these low criterion teachers. If the teachers who have a correlation of .60 are designated as the high criterion group and all others as low criterion, the reliability of their observations can be compared. The mean reliability of the composite for the high criterion teachers was .80 compared to .65 for the low criterion teachers; both calculations were made by the *r to z* transformation. This difference suggests that inconsistent and erratic behavior leading to unreliable observation may characterize some kinds of ineffective teachers. If observers cannot reasonably agree on a description of a teacher, it is even more unlikely that six-year-old children will be able to predict her intentions and attitudes from one time to another either. According to the SCIS criterion, a different kind of ineffectiveness is represented by experienced teacher T, whose observers strongly agreed (.86); however, behavior was related in a substantially negative way to the criterion sort (-.70).

COMPARISONS BETWEEN THE CLASSROOM SORTS

All comparisons between classroom observations are based upon a series of *t* tests performed upon the items of the *Q* sorts. The scores for each teacher are the composites of the individual observer's item placements which are appropriate for a particular analysis; e.g., the composite of <u>all</u> observers who saw the teachers when the contrast is between the experimental and control schools, or the composite of the two observers who saw the teacher the first time contrasted with the composite of the two who

Table 3

SCIS Composite Sort for the Criterion Situation

(items arranged according to scores from highest to lowest)

Scores	Most Characteristic Items (scores from 9.0 to 5.5)
8.5	does not "fish" for predetermined answers
8.2	urges children to try things on their own
8.0	builds on children's disagreements to clarify
7.9	listens to children's responses as a guide to further discussion
7.8	listens to children and respects their points of view
7.8	turns children's questions to the group
7.6	permits children to talk among themselves
7.5	makes opportunities for children to follow their choices
7.5	helps children to discover their own contradictions
7.5	permits relatively free movement among children
7.4	takes into account children's suggestions about procedures
7.3	evokes curiosity in children
7.2	T and children jointly summarize
7.2	individualizes instructions
7.1	has good organization in handling materials
7.0	is excited and interested in helping children make "new" discoveries
7.0	responds to children's mistakes by asking further questions
6.9	is interested in what child is trying to do
6.8	is patient when a child does not catch on
6.5	gives child reasons for doing things
6.3	rewards a child's performance
6.2	reassures children
6.2	reminds children of unanswered questions
6.2	encourages child to try again
6.2	recognizes child's feelings
6.0	responds to children's mistakes by suggesting they think more
5.9	does not need to look at manual
5.6	redirects a child's activity when he misbehaves
5.6	expresses sympathy
5.6	appears certain of herself
	Most Uncharacteristic Items (scores from 1.0 to 4.5)
1.3	shows covert hostility
1.6	allows pupils to ridicule mistakes of classmates
1.8	displays quick temper
1.8	becomes rattled or angry
2.1	lets children know who's boss
2.1	tries to get children to memorize
2.2	is irritated when a child can't find things
2.2	makes evaluative statements comparing one child with another
2.3	uses physical restraint
2.5	"talks down" to children

2.5	tells children they're wrong
2.5	is irritated when a child knocks things over
2.8	emphasizes need for children to be orderly, quiet
2.8	interrupts pupil who's speaking
3.1	does not depart from usual seating arrangements
3.2	responds to children's mistakes by giving correct information
3.2	explanations leave pupils puzzled
3.2	stays in front of classroom
3.3	anxious to complete lesson on time
3.3	makes a point of summarizing
3.5	tries to get children to wait
3.6	T and children show physical affection to each other
3.6	seems confortable with more structured materials and activities
3.7	gives detailed instructions on how to accomplish tasks
3.8	sets limits for behavior
4.4	keeps feelings under tight rein

Items Irrelevant to the SCIS (scores from 5.5 to 4.5)

5.2	readily gives attention to children who need it
5.2	jokes and laughs with children
5.2	seldom corrects child before whole class
5.2	avoids both positive and negative evaluations of child
5.2	approaches disruptive behavior as a problem to be worked out
5.2	gives detailed instructions on the mechanics of procedures
5.1	answers simple questions of fact
4.8	face is animated
4.7	is an active, lively person
4.5	expresses her own feelings

saw her the second time when changes in teacher behavior are being analyzed.

Since each of the 66 different items are possible candidates for distinguishing one group from another, no analysis will be presented unless the total number of significantly differentiating items either reaches or exceeds one of two thresholds: seven or more items must attain the $\leq .10$ probability level or four or more items must attain $\leq .05$ probability level. These rules were established because chance alone can be expected to yield 6.6 significant items (10% of 66) 10% of the time and 3.3 items (5% of 66) 5% of the time. Individual items at the .10 level of probability that are part of an analysis accepted by the previous criterion will be presented, however, because of the exploratory nature of this study. The language of the items in the accompanying tables is telegraphic in form to assist the reader in catching the substance of a set of results.

The Differences Between the Experimental and Control Classrooms

The differences between the teachers in the two study schools were analyzed in three ways by comparing their differences at the time of the first observation; their differences for the second observation; and their differences for all observations combined. The results of these analyses can be seen in Table 4.

Contrasting the characteristic behavior of the two groups of teachers for the overall composite did not achieve the threshold for reporting results that has been established for this study, but the contrast of the behavior for the first observation did, although not impressively. The experimental teachers were seen to be taking a less centralizing and controlling role. The contrast in teacher behavior at the time of the second observation suggested that this same differentiating trend was occurring. Although these results are suggestive, their meagerness dictates caution in assuming that they represent truth. Note that the present method of grouping all experimental teachers and all control teachers together washes out important differences that have been previously shown to exist between individual teachers in both schools.

The Changes Within the Experimental and Control Classrooms

Increasing experience with the curriculum could be expected to produce change in procedures (although it is entirely possible that length of the study and conditions in the experimental school surveyed in this study do not permit an adequate test of change or potential change). The *Q* correlations presented and discussed above showed that some of the teachers made general changes, but this kind of analysis still does not provide information with regard to the specific content of change.

The question is first asked: are there systematic differences in the *Q* items which are distinctive of the experimental teachers in the first observation as compared to the second, and the second as compared to the third? Table 5 shows the results of this comparison.

Table 4

Q-Sort Items which Significantly Differentiated between the Experimental and Control Schools

Item Number	Item	Group A $\bar{X}$	Group A SD	Group B $\bar{X}$	Group B SD	t ratio	N
	Experimental vs. Control - First Observation						
	Items most characteristic of Experimental teachers:						
30**	listens to children	6.42	1.28	4.50	1.22	2.36	6 x 4
	Items most characteristic of Control teachers:						
1***	gives detailed instructions on how to do task	4.75	1.92	3.25	1.19	-3.38	6 x 4
6**	sets limits	5.50	1.67	8.00	1.41	-2.45	6 x 4
12***	reminds children of unanswered questions	4.42	0.58	6.38	1.03	-3.88	6 x 4
65***	gives detailed instructions on mechanics	3.67	1.40	7.25	1.50	-3.86	6 x 4
	Experimental vs Control - Second Observation						
	Items most characteristic of Experimental teachers:						
5***	interested in child's performance	6.33	0.82	4.33	0.29	4.00	6 x 3
19*	approves and rewards child's performance	6.50	0.95	4.67	1.44	2.33	6 x 3
	Items most characteristic of Control teachers:						
23*	keeps feeling under tight rein	3.75	1.37	6.17	2.31	-2.02	6 x 3
25**	interrupts child when he's talking	2.75	0.61	3.67	0.29	-2.40	6 x 3

Note: *$p < .10$; **$p < .05$; ***$p < .01$.

Table 5

Q-Sort Items which Significantly Differentiated between Observations

Item Number	Item	$\bar{X}$	SD	sigma MD	t ratio
	1st vs. 2nd Observations - Experimental School				
	Items most characteristic of first observation:				
3*	is irritated when a child knocks things over	1.33	1.57	0.64	2.08
23*	keeps feelings under a tight rein	1.42	1.36	0.55	2.56
25**	interrupts speaking pupil	1.58	1.32	0.54	2.94
50***	expresses her own feelings	1.17	0.61	0.24	4.72
	Items most characteristic of second observation:				
1*	gives detailed instructions on how to achieve task	-2.50	2.90	1.18	-2.12
12*	reminds children of unanswered questions	-1.50	1.82	0.74	-2.02
16*	makes a point of summarizing	-1.92	1.88	0.77	-2.49
63**	is patient when a child does not catch on	-0.50	0.45	0.18	-2.74
65**	gives detailed instructions on mechanics	-2.92	2.08	0.85	-3.43
	2nd vs. 3rd Observations - Experimental School				
	Items most characteristic of second obesrvation:				
12*	reminds children of unanswered questions	0.92	1.11	0.45	2.02
26***	has good organization in handling materials	1.42	0.74	0.30	4.72

Item Number	Item	$\bar{X}$	SD	sigma MD	*t* ratio
Items most characteristic of third observation:					
3*	is irritated when a child knocks things over	-1.92	1.93	0.79	-2.43
13*	shows covert hostility	-1.42	1.72	0.70	-2.02
20**	explanations leave pupils puzzled	-1.00	0.84	0.34	-2.93
25*	interrupts speaking pupil	-1.50	1.61	0.66	-2.28
51*	tries to get children to wait	-1.33	1.57	0.64	-2.08
64*	displays quick temper	-3.00	2.88	1.18	-2.55

Note: *$p \leq .10$; **$p \leq .05$; ***$p \leq .01$.

The trends between the first and second observations seem to be in the direction expected if experience and training have the desired effect: at the later period the teachers appeared to be less self-concerned and consequently were taking more of a guiding role with regard to children's learning. In the second observation there is indication of some teacher apprehensiveness that the children need very careful instructions. Comparisons between the second and third observations are a different matter: an actual reversal of movement toward the SCIS criterion seems to have occurred; the characteristic third period behavior shows that the teachers are again evidencing more self-supporting behavior. There are several possible interpretations of this reversal: observation occurred late in May, and it is commonly known among school people that fatigue in teachers and disorder in children set in during the closing days of the school year. Moreover, the teachers were all attempting to finish the unit's later lessons, which may be more difficult for an inexperienced teacher to handle, particularly with economically disadvantaged children such as these. Recall, however, that the overall Q results show that several of the experimental classrooms did not follow the trend.

A like analysis compared the first and second observations for the control classroom. Only five items significantly differentiated the periods, suggesting that the differences may be due to chance and that control classrooms were consistent between times. It is always an uncertain matter to assert that negative results have a positive meaning, but, nevertheless, consistency between observations was expected for the control school, and more persuasively, the Q relations show that teachers I and J, who were reliably related, did not change.

The Effects of Varying Degrees of Experience

Although the exploratory nature of this study makes it important to examine the effects of various degrees of experience, the limitations and the tentativeness of the results, which were discussed above, must be kept in mind. The use of the word "experienced" in this context is relative. The tenure of the non-project teachers in the curriculum ranged from one to four years with an average of 2.4 years, whereas the experimental teachers were comparatively "inexperienced." Two analyses are possible: the experienced teachers are compared with the inexperienced, and the experienced criterion-high teachers are compared with the inexperienced criterion-high teachers. The third analysis which would be a logical completion of this sequence is not possible, since two of the three low criterion, experienced teachers have had only a few more months' experience than the experimental school teachers.

The results for the first analysis (comparing all experienced with inexperienced teachers) are shown in Table 6. Their import is clear: during science instruction the experienced teachers were more involved in cognitive activities and less involved in personalized interaction with children. This finding is consistent with SCIS's aims of providing curricular structure that is so involving to children that their curiosity, unfettered by boredom and rebellion against adult coerciveness, leads them to cognitive engagement.

Table 6

Q-Sort Items which Significantly Differentiated between Experienced and Inexperienced Teachers

Item Number	Item	Group A $\bar{X}$	Group A SD	Group B $\bar{X}$	Group B SD	*t* ratio	N
	All Experienced vs. All Inexperienced Teachers (SCIS)						
	Items most characteristic of experienced teachers:						
24*	redirects child's activity when he misbehaves	3.89	1.23	5.25	1.60	-1.78	6 x 10
46*	face is animated	6.56	1.14	7.55	0.93	-1.91	6 x 10
47*	approaches disruptive behavior as problem to be worked out	3.36	1.22	4.75	1.48	-1.94	6 x 10
55**	builds on disagreement to clarify	4.89	0.49	7.00	1.96	-2.56	6 x 10
56*	T and children jointly summarize	3.69	0.63	5.05	1.57	-2.00	6 x 10
	Items most characteristic of inexperienced teachers:						
48**	uses physical restraint	5.64	1.68	3.65	1.36	2.61	6 x 10
52*	T and children show physical affection	5.53	1.54	4.05	1.21	2.14	6 x 10
	Experienced Criterion vs. Inexperienced Criterion Teachers (Experimental School)						
	Items most characteristic of experienced criterion teachers:						
7*	tries to get children to memorize	3.92	0.49	3.22	0.35	2.16	6 x 3
27*	evokes curiosity in children	7.25	0.82	5.89	1.02	2.18	6 x 3
28*	excited and interested in helping children make new discoveries	8.17	1.21	6.50	0.73	2.15	6 x 3
30*	listens to children and respects their points of view	8.00	0.63	7.00	0.39	2.33	6 x 3

Item Number	Item	Group A $\bar{X}$	Group A SD	Group B $\bar{X}$	Group B SD	t ratio	N
32*	helps children discover their own contradiction	6.50	1.22	5.00	0.73	1.92	6 x 3
38*	appears certain of herself	7.33	0.75	6.33	0.33	2.14	6 x 3
41**	listens to children's responses as a guide to further discussion	7.58	1.16	5.56	0.54	2.81	6 x 3
55***	builds on disagreement to clarify	7.92	1.24	5.17	0.44	3.62	6 x 3
60*	turns questions to group	7.58	0.86	6.50	0.60	1.93	6 x 3
	Items most characteristic of inexperienced criterion teachers:						
6**	sets limits for behavior	4.08	1.24	6.44	0.25	-3.16	6 x 3
9***	lets children know who's boss	1.83	0.52	4.28	1.02	-4.96	6 x 3
25*	interrupts speaking pupil	2.33	0.68	3.28	0.67	-1.96	6 x 3
34*	makes evaluative statements comparing one child to another	2.42	0.66	3.39	0.35	-2.32	6 x 3
58*	becomes rattled or angry	2.17	0.75	3.94	1.84	-2.14	6 x 3
61**	anxious to complete lesson on time	2.42	0.20	3.72	1.25	-2.67	6 x 3
62**	answers simple questions of fact	4.58	0.20	4.94	0.10	-2.84	6 x 3
64*	displays quick temper	2.00	0.89	3.17	0.73	-1.94	6 x 3

Note: *$p \leq .10$; **$p \leq .05$; ***$p \leq .01$

The difference is even clearer in the next analysis where experienced and inexperienced teachers, both of whom satisfy the SCIS criterion, are compared with each other. (Note that these are comparative results. The inexperienced criterion teachers were distinguishable from their colleagues in the same school by 16 items and these include many which are comparatively uncharacteristic of them within this context. These results are not presented here, since separating teachers into groups on the basis of their correlation with the criterion sort and then anlyzing their differences on the same sort is a circular process which predetermines the findings.)

In general, the results of these analyses indicate that the more experience with SCIS the teachers have, the more likely they are to take on criterion behaviors, and that these behaviors are particularly noteworthy in their intent to facilitate and incite children's cognitive involvement and curiosities. The results also indicate that the experienced teacher becomes more of a guide and an attentive listener, at least during science instruction, and that she does not need to react to children during this time in a controlling and effective, personalized way. Moreover, experienced teachers are not only more certain of themselves, but are also able to listen more attentively to the children.

THE CHILDREN

Six first-grade classrooms in the experimental school and four in the control school constituted the sample of 167 children who were both pretested and posttested. Two classes in the experimental school and one in the control school were not included because they represented special situations for children with severe learning problems, emotional or otherwise. One other class in the control school was pretested but was not studied further after their original teacher left the school. Aside from these exceptions, the study sample includes all first grade children in both schools.

Table 7 presents a number of different kinds of background data for the children. In general the information shows that the children in this sample are from homes where the main wage earner is a skilled worker and that a comparatively large number of the mothers work full-time or part-time. Comparison of the two groups showed that there were no significant differences in their age, the occupational level of the fathers or the mothers, the size of their families, or the subjects' birth order (these background variables were also analyzed for the sexes separately, and no significant differences were found). There are a few more children from bilingual homes in the control school.

All the testing of the children was done by the same principal examiner who was an experienced, primary grade teacher; she was assisted by another person who had also had teaching experience, as well as a third for some parts of the test administration which involved particularly complicated logistics. The decision to have this many adults in the room was made to prevent disciplinary situations. The classes' regular teachers

Table 7

Description of the Child Sample

				Parental occupation[2]			Background		I.Q.
	Pretest N	Posttest N	Core Sample % of Pretest	Father $\bar{X}$	Mother $\bar{X}$	Mother working[3]	Family size $\bar{X}$	Bilingual family	$\bar{X}$
Experimental									
Boys	57	52	91%	4.12	4.26		3.31		
Girls	61	55	90%	4.55	3.96		3.34		
Total	118	107	91%	4.36	4.11	45%	3.32	6%	94.44
Control[1]									
Boys	40	33	82%	4.28	4.50		3.58		none
Girls	30	27	90%	4.43	3.70		3.15		available
Total	70	60	86%	4.34	4.10	31%	3.38	13%	
Overall	188	167	89%						

Note.--[1]An additional control school class of 13 children was pretested but not posttested since their original teacher left the school.

[2]Occupational level was classified according to the Edwards' scale: (1) professional, semiprofessional; (2) promoters, managers, officials; (3) clerks and small proprietors; (4) skilled workers; (5) semiskilled workers; (6) unskilled workers.

[3]Includes full time and part time.

were not present during pretesting but were invited to be there during posttesting if they wanted to; however, none of them actually stayed in the room more than a few minutes.

The discussion will now turn to a presentation of the results; each assessment method will be described in turn, and the results of the various comparisons will immediately follow. It will be possible to make more analyses of the children's data than the teachers' because larger numbers of children are, of course, available. The following seven contrasts were made and the results will be reported in this order: (1) Control vs. Experimental school (boys and girls combined -N's = 107 vs. 60); (2) Control vs. Experimental schools (boys -N's = 53 vs. 32); (3) Control vs. Experimental school (girls -N's = 54 vs. 28); (4) High-Criterion Experimental vs. Control classes (N's = 58 vs. 13); (5) Low-Criterion Experimental vs. Control classes (N's = 49 vs. 29); (6) High-Criterion classes vs. Low-Criterion classes (experimental school -N's = 58 vs. 49) and (7) High-Criterion classes vs. Low-Criterion classes (control school -N's = 13 vs. 29).

All analyses are done by the *t* test for comparisons of changes (McNemar, 1962). This procedure means that the scores shown in the subsequent tables are change scores. The reader who is interested in the absolute levels of pretest and posttest scores can find them in Appendix D for all the analyses that are reported below.

These indices are secured by subtracting the pretest scores from the posttest scores so that a change score with a positive sign indicates a gain whereas one with a negative sign indicates a loss. Change scores involve some advantages, disadvantages, and complexities. They are particularly vulnerable to being distorted by chance fluctuations in subjects' responses since two scores, each with its own uncertainty, must be used to produce one change score. Their effectiveness is also markedly diminished if the range of difficulty presented by measurement is not sufficient. A high score obtained at pretest because of a curtailed range of difficulty can sometimes be followed by a slightly lower score at posttest because of a random fluctuation in a subject's attention. If this occurs, the subject's change score registers as a loss even when in reality he may have made some gain. However, change scores do provide a kind of information that is not summarized as directly by other kinds of statistics. Within some limits, change scores have the advantage of exerting some control over differences in the level of various individual's initial scores since a score of "2" at pretest and a score of "7" at posttest results in a change score of "5", just as an initial score of "10" and a posttest score of "15" results in a change score of "5".

The SCIS Tasks

Three assessment tasks were designed by the SCIS staff: the first required children to select six objects from an array, according to various properties of color, size, and weight (Selection by Property); the second required the children to sort an array of screws according to their materials

of plastic, brass, steel and aluminum (Sorting by Material); the third task required that children place washers in seriate order according to their size (Seriation). A scoring procedure was devised to give each child credit for his proficiency in accomplishing each of these tasks.

Although some pilot runs were made with children in other schools to improve the format and instructions for these tasks, the pressure of time did not permit adequate pretesting. As it turns out, the task of Sorting by Material was far too easy a task for these children. Most children failed to distinguish between aluminum and steel, but if this requirement is set aside as representing a sharp discontinuity in the difficulty and scoring of the task, 32% of the total sample had a perfect or near perfect (one less than perfect) score on the pretest and 66% on the posttest. Therefore, this task is not suited for studying gains since it did not permit one-third of the children to register their gains. It can be noted, however, that the percentage of experimental children who showed improvement by achieving near perfect scores at posttest (35%) exceeded the percentage of control children who did (18%).

A few children also attained near perfect scores for the other two tests (Selection by Property: experimental school, 13%, and control school, 14%; Seriation: experimental school, 8% and control school, 15%). This occurrence suggests that these two tasks would also be better assessment techniques if they were made more difficult. Since the "ceiling effects" for these two are smaller, results of their analysis will be reported.

The information relevant to these anlyses is to be found in Table 8. The task of selecting objects according to their properties clearly does not differentiate between children from the two schools or between children within either of the schools when they are grouped according to their teachers' relationships to the criterion sort. Moreover, examination of the direction of the t ratios for all comparisons does not show a trend favoring one group or another.

Quite a different set of results are shown with the Seriation test; here three sets of results are significant in the expected direction and all but one of the results which do not achieve significance indicate a trend favoring the experimental school children and the high criterion classrooms.

The explanation of these two results--a confused set from one test and an understandable set from another--suggests the possibility that the Selection by Property task may have a low reliability, at least when it is administered as a group test. It is apparent that performance in the Seriation task requires not only that the child have mastery of an abstract concept, but that he also have the capability to operate on it; whereas success in the Selection by Property test, at least in a group situation, may be more related to paying attention and following instructions (providing the child can differentiate among white, gold, black, silver, smallest

Table 8

Comparison of Gains for SCIS Tasks

	Selection of Property					Seriation				
	Experimental $\bar{X}$	Experimental SD	Control $\bar{X}$	Control SD	t ratio	Experimental $\bar{X}$	Experimental SD	Control $\bar{X}$	Control SD	t ratio
Experimental vs. Control										
All (107 x 60)	1.34	3.32	1.40	3.82	-.11	1.46	3.13	1.05	3.25	.77
Boys (53 x 32)	1.37	3.14	1.30	4.43	.09	1.59	2.93	1.13	3.29	.65
Girls (54 x 28)	1.31	3.51	1.52	3.07	-.26	1.33	3.34	.96	3.26	.47
High-Criterion (58 x 13)	1.54	3.33	.83	3.04	.67	2.41	3.29	-.08	3.68	2.33**
Low-Criterion (49 x 29)	1.12	3.33	1.96	4.66	-.91	.37	2.57	1.63	2.63	-2.03**
High- vs. Low-Criterion										
Experimental (58 x 49)	1.54	3.33	1.12	3.33	.63	2.41	3.29	.37	2.57	3.51***
Control (13 x 29)	.83	3.04	1.96	4.66	-.77	-.08	3.68	1.63	2.63	-1.66

Note: *$p \leq .10$; **$p \leq .05$; ***$p \leq .01$

and lightest; the test instructions supplied him with the labels for screw, washer, and nut).

Attending is not an unimportant consideration for the SCIS program, but it is also clear that it is not unique with this curriculum. Furthermore, one of the manifest advantages of the SCIS program is that its structure and ultimate success is not contingent upon children's giving prolonged attention to a single, central classroom activity, but this test, more than the Seriation, requires that children sustain their attention through a series of six instructions. Therefore, it is not a good model of the positive gains for children that can be expected from a SCIS experience. This explanation of the present findings is further supported by the fact that the girl's pretest scores for Selection by Property were somewhat higher ($p \leq .10$) than the boys. It is, of course, a common observation that first-grade girls are usually more socially compliant and pay better attention than first-grade boys. As it was, it appears the present findings are due to random fluctuations in attention and consequent low reliability, rather than to gains or losses in the children's understandings.

Further, it should be noted in Table 8 that aside from the minimal loss of the children in the High-Criterion classes from the control school, all other means indicate gains by their positive signs. This result is due to at least two factors: practice effects in taking the tests and/or the growth in general capability of these six- and seven-year-old children during the five-month period.

The General Cognitive Tasks

All of the measuring devices used were taken or adapted from the recently published Let's Look at First Graders (Board of Education of the City of New York, 1965). This small curriculum and set of evaluation procedures were developed jointly by Educational Testing Service and the Board of Education of New York City to foster intellectual development in young children; its underlying conceptualizations are heavily influenced by Piaget. Four tasks were chosen: two series of booklets to evaluate (1) logical reasoning and (2) conceptualizations of shapes and forms; and two tasks to evaluate (3) understanding of gravity and (4) conservation. As it turned out, the Shapes and Forms tasks also had too low a ceiling to be useful in this investigation of gains: 53% of the children had perfect or near perfect scores at pretesting (there is probably good reason to think that Berkeley children may be generally more advanced than New York City children, but unfortunately this is an after-thought.

Logical Reasoning presents a series of 30 problems; the children are requested to indicate by means of "x's" which of a continuing series is analogous or properly grouped with a specimen (according to color, function, size, and the like).

For the Understanding Gravity tasks, the examiners showed the children two Mason jars, one empty and the other half-filled with colored water. Each child had an answer sheet which showed empty Mason jars in three positions: upright, at a 45^{0} angle, and horizontal. The children were asked to color in the position that the water would take if the jar were tipped; the empty jar was used to show graphically the two angulations.

During the administration of the Conservation task the two examiners stood in front of the room and asked the children to count as 20 large beads were simultaneously dropped into two jars--a broad flat one and a tall thin one. The latter jar gave a gross impression that it contained more beads, but the children's counting in unison provided all of them with the same information, that is, each jar had the same number of beads in it. This typical Piaget Conservation task was meant to assess the children's ability to maintain the primacy of verified information when it was seemingly contradicted by visual stimuli. Again, each child had an answer sheet where he was to indicate his choice as to which jar had the most beads or whether both jars had the same number.

No significant differences in gains resulted from the analysis of the Logical Reasoning scores, although the direction of results show that the experimental school children made the greater gains (cf. Table 9). There is a disconcerting reversal when the criterion classrooms within each school are compared with their non-criterion counterparts but the differences are not significant. Another insufficiency in testing needs to be noted: only 9% of the children had perfect or near perfect scores at pretesting, whereas 41% did at posttesting. This pattern suggests that this task is evidently sensitive to practice effects, which does not seem plausible since there were many questions and none was particularly vivid or highly reactive to growth gains generally made by first graders as they mature and learn during the school year. Since this large proportion of the children pressed against the test ceiling by May, this instrument obviously does not permit an adequate test of gains.

Analysis of the differences in gains with regard to Understanding Gravity shows that the High-Criterion experimental classrooms gained significantly more than the High-Criterion control and the Low-Criterion experimental. The trends and results for the Conservation task are somewhat similar. Greater proportions of the experimental children (particularly those in the High-Criterion classrooms) conserved at the time of posttesting than did the control children or those in Low-Criterion experimental classrooms. Again, the Low-Criterion children in the control school outperformed their schoolmates in the High-Criterion.

In general, then, the SCIS children, particularly those in the criterion classrooms, show a nonsignificant, but positive trend toward greater cognitive attainment. This is most evident with Understanding Gravity and Conservation.

Table 9

Comparisons of Gains for General Cognitive Tasks

	Logical Reasoning					Understanding Gravity					Conservation		
	Experimental $\bar{X}$	Experimental SD	Control $\bar{X}$	Control SD	t ratio	Experimental $\bar{X}$	Experimental SD	Control $\bar{X}$	Control SD	t ratio	Diff	SE	CR[1]
Experimental vs. Control													
Boys and Girls	3.19	4.18	2.07	4.28	1.63	.74	1.69	.65	1.85	.30	.13	.09	1.39
Boys	2.90	3.37	2.31	4.59	.67	.77	1.78	1.07	2.03	-.68	.14	.14	1.06
Girls	3.46	4.84	1.77	3.94	1.55	.70	1.61	.19	1.52	1.37	.11	.12	.92
High-Criterion	2.58	3.16	.92	3.12	1.66	1.04	1.66	.15	1.99	1.67*	.46	.20	2.34**
Low-Criterion	3.92	5.08	2.04	4.30	1.65	.39	1.68	.58	1.47	-.48	-.07	.12	-.57
High- vs. Low-Criterion													
Experimental	2.58	3.16	3.92	5.08	-1.65	1.04	1.66	.39	1.68	1.99**	.10	.11	.96
Control	.92	3.12	2.04	4.30	-.81	.15	1.99	.58	1.47	-.75	-.42	.20	-2.09**

Note: *$p \leq .10$; **$p \leq .05$; ***$p \leq .01$

[1] Calculated by a method developed by Hovland et al., 1949, pp. 337-40.

Locus of Control

The general idea of Locus of Control was first suggested by Rotter (cf. Battle and Rotter, 1963). However, the present task represents a considerable departure from Rotter's conceptualization.[3] In the present study the child was asked to choose a picture from a pair, one of which represented himself. Each child had his own set of nine paired pictures and an envelope in which he was to place his choices: the two examiners held up enlarged members of the pairs for the children to see as they were told a story for each situation. An example story would be: "Here's a boy who gets an extra piece of cake, does he get it because no one else wants it or because he earned it helping his mother?" The pairs of pictures were drawn to illustrate each alternative. The alternatives offered the child a choice as to whether he sees himself as instrumental in determining his own fate (internal control; in the example, "He earned it") or whether he sees his fate as being fortuitously determined for him (external control; in the example, "No one else wanted it"). Thus, a score for total internal control is generated. In addition, various situations have a positive or negative flavor which permit two other sub-scores: positive internal and negative internal. The example depicted a positive occurrence, but a negative one would involve a boy dropping a dish because he didn't hang on tightly enough (or because someone bumped into him--negative external). The alternative, negative or positive, which was mentioned first was systematically varied from one situation to the next to counteract any tendency that individuals might have to become fixed on the first possibility.

Since the SCIS curriculum urges that children be encouraged to take more responsibility for drawing their own conclusions and for learning in general, and because the procedures of the curriculum promote that end, it follows that children who have had this experience should increase their sense of being able to determine their own destiny. Table 10 shows that this is the case; all comparisons of experimental with control children show that the former increased the number of their choices of internal self-determinations at posttesting and that this movement was particularly true of boys and the children in the High-Criterion experimental classrooms. The trend was most marked for positive internal situations but occurred as well for the negative internal situations with the children in the High-Criterion classrooms.

Although this technique does not have a history of validation and replicated findings, the present results and its face validity make it appear that it may be useful in further assessment of the effects of this curriculum and other such "new" curricula which intend to increase the autonomy of children as learners.

---

[3]Frances Zavola of the University of California designed the orginal task for the Lenrow Preschool Project; the present adaptation for group testing was done by Peter Lenrow and Norma Haan.

Table 10

Comparisons of Changes toward Internal Locus of Control

	Total Internal					Positive Internal					Negative Internal				
	Experimental		Control		*t*	Experimental		Control		*t*	Experimental		Control		*t*
	X	SD	X	SD	ratio	X	SD	X	SD	ratio	X	SD	X	SD	ratio
Experimental vs. Control															
Boys & Girls	.02	1.57	- .70	2.15	2.41**	.43	1.24	-.21	1.47	2.89***	-.34	1.35	- .46	1.67	.50
Boys	.12	1.47	-1.00	1.80	3.03***	.40	1.18	-.53	1.31	3.29***	-.24	1.21	- .41	1.78	.53
Girls	-.08	1.68	- .37	2.48	.61	.45	1.32	.15	1.59	.90	-.45	1.47	- .52	1.58	.19
High-Criterion	.04	1.41	-1.75	1.66	3.85***	.41	1.30	-.50	.90	2.30**	-.39	1.15	-1.25	1.96	2.04**
Low-Criterion	-.00	1.77	- .41	2.24	.85	.44	1.18	-.07	1.49	1.63	-.28	1.56	- .37	1.57	.23
High- vs. Low-Criterion															
Experimental	.04	1.41	.00	1.77	.11	.41	1.30	.44	1.18	-.13	- .39	1.15	- .28	1.56	-.41
Control	-1.75	1.66	-.41	2.24	-1.86*	-.50	.90	-.07	1.49	-.91	-1.25	1.96	- .37	1.57	-1.49

Note: *p≤.10; **p≤.05; ***p≤.01

The Other Attitudinal Measures

Since the SCIS program intends that children's curiosity be stimulated, a task was adapted from one used in a study of preschool children by Medinnus and Love (1965). A small black box with a colorful curtain for a front plate and a toy hidden inside it was placed on a table in front of the room. Each child had an answer sheet to indicate his play preferences between paired choices of toys: the one hidden in the box or the one displayed by the examiner (the children were shown the hidden toy after they had completed all their choices). The toys displayed were selected to be relatively uninteresting to six- and seven-year-old children, e.g., modeling clay, coloring book and crayons, etc. There was insufficient time before pretesting began to establish whether these were uninteresting to children but examination of the preference splits for individual toys established the fact that none of them was so attractive that more than 50% of the children chose it both times. However, it may be questionable to assume that a task involving toys of any kind is a bland matter to a group of relatively economically disadvantaged children. Toys, known or unknown, may be a matter of such consequence to them that their choices were not made on the cognitive level, as it was intended. It is the writer's belief after observing the children's high involvement with the task that it elicited, instead, fantasies of fulfilling unsatisfied needs.

Another attitudinal test was adapted from the work of Kuno Beller (1964) who showed that preschool children's dependency strivings could be measured by the discrepancy which existed between their indications of their own height and that of adults. In this study the children were given a blank piece of paper, asked to fold it in half lengthwise, and then starting at the floor (the bottom of the paper) to draw a line as to how tall each of them was and then another for how tall their teacher was.

Scores were generated for each of these tests: the score for Unknown Toys was a simple count of the number of choices that the children made for the hidden toy instead of the displayed one; the score for child/adult height was obtained by measuring the children's indications of their height and their teacher's height and establishing a ratio between the two. The comparisons of the gains made by the various groups is shown in Table 11.

The results for Unknown Toys do not give support for the conceptualizations that guided the selection and the design of this test: instead, the evidence actually suggests a mild trend in the opposite direction. It becomes a significant difference within the experimental school when the children in the High-Criterion classrooms are contrasted with the Low-Criterion classes. Another aspect of high affective load of this test needs to be noted: the toy that was hidden during pretesting was a pair of skates and it clearly was an attractive item to the children, who responded with "ooh's" and "aah's." Moreover, this experience was sufficiency vivid that many of these young subjects reminded the examiners five months later that skates had been hidden the first time. Thus, choosing the unknown toy on the second round of testing was probably an indicant of a more intense affective reaction than of cognitive curiosity.

Table 11

Comparisons of Changes toward Choices of Unknown Toys and Equability in Child-Adult Height

	Unknown Toys					Child/Adult Height Ratio				
	Experimental		Control		*t*	Experimental		Control		*t*
	X	SD	X	SD	ratio	X	SD	X	SD	ratio
Experimental vs. Control										
Boys & Girls	1.22	4.22	2.07	3.93	-1.26	.36	29.67	-2.71	22.72	.66
Boys	.94	4.07	1.97	3.85	-1.12	1.65	31.79	-6.93	27.75	1.19
Girls	1.49	4.39	2.19	4.10	-.68	-.91	27.68	1.84	14.86	-.46
High-Criterion	.35	3.69	1.08	4.21	-.62	4.36	24.33	-6.42	21.84	1.42
Low-Criterion	2.22	4.60	1.69	3.79	.51	-4.20	34.48	-1.23	19.45	-.38
High- vs. Low-Criterion										
Experimental	.35	3.69	2.22	4.60	-2.33**	4.36	24.33	-4.20	34.48	1.48
Control	1.08	4.21	1.69	3.79	.46	-6.42	21.84	-1.23	19.45	.71

Note: *$p \leq .10$; **$p \leq .05$; ***$p \leq .01$.

The results for the Child/Adult Height Ratio are not impressive; there is a slight trend toward the boys in the experimental school and toward children in High-Criterion classrooms in the experimental school closing the gap between themselves and adults during the five months' experience but it cannot seriously be considered a result.

Children's Questions with Regard to the Curious Objects

Some investigations have used the quantity or quality of children's questions about unusual situations to index the degree of their curiosity (cf. Haan, 1967, for a review of these studies). If individual measures of questioning were desired, children would have to be seen one at a time or be capable of expressing themselves fluently with paper and pencil. Neither of these conditions could be realized in the present situation. Therefore, two Curious Objects were devised, one for pretesting and another for posttesting;* they were shown to the group as a whole with the announcement that the children would probably like to ask questions about it and that each child could have one turn. The second examiner recorded the children's questions. The questions for each class provided the basis of group measures rather than ones for individual children. This assessment procedure results in a group contagion as one child's question tends to influence another's. However, this phenomenon in itself was of interest to the assessment since it was assumed that the SCIS experience should encourage children to pose individuated questions rather than dependently and compliantly following other's lead.

The questions were typed without identifying information and submitted for categorization to two psychologists who were not involved in this project and had no knowledge of its underlying hypotheses. A classification scheme of questions had been developed by the writer and project assistants which was partly derived from the present questions themselves, but also drew upon work by Fahey (1942). Table 12 shows the scheme that was developed and the proportions of questions for the two testings that were classified in different categories for the entire sample. Precise agreement was obtained between the two judges on the classification of 89% of the pretesting and 80% of the posttesting questions. Another person independently judged all the questions and this third, independent opinion permitted an additional 9% of the controversial questions for the pretest and 17% for the posttest to be included by a single majority vote of two out of three judgments.

Since variety in the classification of the groups' questions reflects their capabilities in shifting sets and individualizing their questions, the data were analysed in this regard with the sixth unclassifiable category excluded. Table 13 shows that an average of close to three categories was used by all classes for both the pretest and posttest. When the shifts in the number of categories used is examined, there is

---

*Fiona de Groot was the designer and constructor of these two marvelously intricate and very curious objects.

Table 12

Classification for Questions About Curious Objects and Percentage of Questions Classified for Pretesting and Posttesting

	Pretest	Posttest
1. Identification question Questions which ask "what is it?" or "what is some part of it?"	19%	26%
2. Construction or origin question Questions which ask "how did you make it?" or "where did you get it?"	21%	12%
3. Object's use or performance Questions about the object's function - "can you do something with it?"; how the object can behave - "can it jump?"	35%	14%
4. Guess at identification A guess is made as to what it is, or what some property or part is - "is it a _____?", "is it made of _____?" (Brokenness should be considered a property.)	15%	31%
5. Complex, imaginative question Question which indicates complexity, imaginativeness, or subtlety of thought.	1%	10%
6. Unclassifiable Garbled questions, nonsense or non-feasible questions, and those on which judges disagreed.	8%	7%

Table 13

Numbers of Categories Used to Classify

Questions About the Curious Objects

(by Classroom)

	Pretest	Posttest	Shift
Experimental			
High-Criterion	3	4	+1
High-Criterion	1	4	+3
High-Criterion	3	3	0
		Total	4
Low-Criterion	4	4	0
Low-Criterion	5	2	-3
Low-Criterion	3	2	-1
		Total	-4
Control			
High-Criterion	3	4	+1
		Total	+1
Low-Criterion	3	3	0
Low-Criterion	3	2	-1
		Total	-1

no salient difference between the two schools, but clear differences do occur between the three SCIS High-Criterion and the three Low-Criterion classes: the High-Criterion classes used four more categories at posttesting than they did at pretesting, whereas the Low-Criterion classes used four less.

Before the reader's attention is turned to the shifts within the specific categories, a few more suppositions and descriptions about the Curious Objects need to be given. It seemed to this observer and the examiners that children were rather perplexed by the first Curious Object; not many children trust themselves to challenge or suspect adult sensibilities or correctness, and most children at pretest seemed to be trying to make some sense out of this nonsensical object, presented by two unfamiliar and sober adults. This effort seems to be reflected in the entire samples' high proportion of use and performance questions (35%), which drops radically at posttesting (14%), and the similarly high construction and origin questions (21%), which are also followed by a posttest drop (12%). More to the spirit of the occasion and discerning of adult culpability are the guesses at identification which involved gamesmanship (is it a prize of some sort?) or complex imaginative questions (is it a lady with a purse, sitting down or maybe walking?).

Turning now to Table 14, the reader will note that all groups decreased in their efforts to discern the use of the objects and in their attempts to establish its origin and construction. It seems that it had become generally clear that the object couldn't have a use and that somebody had made it. On the other hand, guesses at identification and complex imaginative questions increased, as the children came to recognize that this was a playful bit of adult nonsense. However, this pattern is most marked for the experimental sample and most particularly for the experimental High-Criterion classes.

Comparisons of Children of Varying Socio-economic Status within the Experimental, High-Criterion Classrooms

Another examination was made with regard to the curriculum's effects upon children whose backgrounds might be assumed to provide varying degrees of intellectual stimulation. In order to control the effects of curricular experience and the impact of their teacher's criterion-related behavior only the experimental children in High-Criterion classrooms were studied. The children from families whose main wage earner is a white-collar worker were contrasted with those of a blue-collar background. This analysis asks the question whether the SCIS-related gains of children from blue-collar background are less, equal, or more than those of white-collar background? (Another analysis would also have been of interest: do blue-collar children in experimental, High-Criterion classrooms gain more than those in the control High-Criterion classrooms? However, the small number of children in the control High-Criterion classes and the uncertain socio-economic data available for some of these children made this analysis infeasible.)

Seventeen children were identified as being of white-collar and 53 of blue-collar background. Five additional children were not included in

Table 14

Classification Shifts Between Pretest and Posttest

in Percentages of Questions Asked

By Various Samples

	Entire Sample	All Experimental	All Control	Experimental High Criterion	Experimental Low Criterion	Control High Criterion	Control Low Criterion
Identification	+7%	+4	+10%	+11%	0%	-10%	+30%
Construction	-9	-13	+1	-17	-9	0	0
Use, performance	-21	-23	-18	-33	-16	-20	-28
Identification guess	+16	+19	+8	+19	+23	+7	0
Complex, imaginative	+9	+14	0	+25	-2	+9	0
Unclassifiable	-1	-2	-1	-4	+3	+8	-2

the analyses because one of the parents was identified as a blue-collar worker and the other as a white collar worker. The group's scores for both the pretest and the posttest and the change between these two times were compared on all of the tests that have been previously reported (cf. Appendix E for the supporting figures). None of the comparisons of change scores reached significance. Examination of the analyses for the absolute scores does not suggest that there were important differences in the initial status of the children from blue-collar and white-collar families. There was a pretest difference of significance: Negative Internal Control was higher for the blue-collar children (t = 2.06, $p \leq .05$) and two posttest differences: white-collar children conserved in greater numbers (x^2 = 2.95, $p < .10$) and the blue-collar children were higher for Total Interal Control (t = 2.95, $p \leq .05$). It appears then that the curriculum had generally comparable effects on both groups. Their initial differences were small; they made rather similar gains and in the end a slightly greater proportion of the white-collar children were able to conserve and more blue-collar children felt they were able to determine their own fate.

Discussion

An evaluation of the results of this exploratory investigation must take account of the limitations in design and measurement and the constraints imposed by external conditions. It clearly has not been a definitive or sensitive test of the effects of the SCIS program on either children or teachers. A detailed attempt was made, however, to study the complex processes and outcomes that seemed to be reasonably attributable to the intentions, structure and procedures of this curriculum. The reader of the curriculum guide for the Material Objects unit must be impressed that SCIS aspires to build a way of thinking about nature and to facilitate children's causal thinking.

A study that would evaluate this curriculum must then try to model its design on the same level of complexity and sophistication that is represented in the curriculum. However, there is a real, and presently insurmountable, obstacle: behavioral science has not yet provided the means of assessing many of the most significant educational outcomes with imaginativeness or with certain reliability and validity. This deficiency is even more true if the subjects are very young children.

It is also widely and sensibly recognized that the teacher, with her power to flavor and determine experiences of young children, is the single most important influence in the classroom, yet very few ways of objectively and reliably describing her behavior in appropriate detail are available. There are some noteworthy exceptions--e.g., Flanders' Interaction Analysis (1963), but most of these instruments rely heavily upon teachers' verbalizations so that they are minimally useful in evaluating curricula which would have teachers talk less and children think more. Moreover, the scores which result from most extant methods tend to be relatively global and presumably reflect teacher effectiveness in general, whereas this curriculum is fairly circumscribed in its

recommendation of a role it would have teachers take during science instruction. Presumably, its authors would be hospitable to teachers taking somewhat different roles when dealing with other aspects of the children's growth.

Since this was the state of affairs--a curriculum which expects more than factual learning and attempts to place responsibility for learning upon the children themselves, a dearth of measures relevant to these expected outcomes, and no means of describing teacher behavior specifically related to this curriculum--the present study necessarily became exploratory and developmental. Nevertheless, it produced provisional answers to some questions and seems to have raised some others.

An important aspect of this investigation was the generation of a detailed description of the kinds of classroom interactions initiated by teachers which the SCIS staff believes are important to children's achieving the goals of the curriculum. The SCIS members had a high degree of consensus among themselves with regard to this criterion description, so that it became possible to take the average score of the 12 members who did the *Q* sorting as the frame of reference for studying the effects upon teachers, and in turn, their effects upon children.

The results indicated that teachers who had had more experience with the program focussed to a greater extent on children's thinking processes and took more of a facilitative (as opposed to a directive role) during science periods than did those who were less experienced, even when the less experienced matched the SCIS criterion behavior rather well. This finding does not establish a causal connection; it would have been necessary for the same teachers to be observed over time, and the five months' study of the experimental school involved too few teachers and too short a time to permit the present investigation a proper examination of the differential effects of experience. But the results indicate that some of them did adopt more criterion behaviors. Taken together, these two findings suggest that a cause and effect relationship may exist. A curriculum which expects important shifts in both teacher and child behavior will obviously take a longer time to do its work, particularly with classroom groups that established a very different modus operandi earlier in the school year. Some of the teachers in the project schools matched the SCIS criterion before they had the experience, some in the experimental school changed in the direction of the criterion, but others were not affected, at least during the time encompassed by this study.

There is some evidence that teachers who closely matched the SCIS criterion role presented clearer impressions to the observers. It was suggested that observers might have clearer judgments of effective teachers since pupils, like the adult watchers, would have greater confidence in being able to predict their teachers' attitudes and expectancies. Within this study, clarity and consistency in a teacher would probably be related to her sense of purposes rather than to a vivid display of herself as a central power figure.

The results for the children are built upon a number of assessment procedures, specifically designed for this study or adapted from other studies which were not comparable. This necessity for invention was unfortunate because the present results cannot now be specifically evaluated within frames of reference. The reader should note, however, that educational change, fostered by the "new" curricula, will inevitably lead to new goals; therefore, subsequent lags in the proficiency and cogency of assessment are inevitable.

The present results show that, in the experimental school, teachers' criterion-related behaviors were positively associated with the gains made by their students. The possibility can be raised that both the teachers' behavior and their students' gains may have occurred because these children might have been more intellectually capable. These measures were not related to I.Q. (the highest correlation was for Logical Reasoning, .43 Selection by Property was .16: Seriation, .22). This alternative explanation might occur to some readers since most of the mean, pretest scores for the children in the SCIS Criterion classrooms are higher than those for the Low-Criterion groups. (Note, however, that the opposite pattern occurred in the control school with the High-Criterion group also having higher pretest scores but fewer gains in some instances. The following discussion will suggest that this pattern can be expected to occur more frequently.) Since an adequate I.Q. measure was not available, the most direct way of examining whether more intelligent children in the sample tended to have High-Criterion teachers was not possible; an equally viable explanation would suggest that the presence of capable children might have permitted these teachers to behave according to the criterion.

Even if either of these alternative explanations were true, neither would sufficiently explain the child gains in the experimental school nor modify the findings except within a narrow range of I.Q., since the average I.Q. for the children who had such scores was 94 and the highest reported was 126. The I.Q. measurement available, of course, was only a group test, a situation which does not promote any great confidence in the validity of the I.Q. data. Change scores, although generally less reliable than absolute scores, are not strongly affected by differences in initial level except where scores are high and the upper reaches of measurement limited, as they were to some extent in this study. This phenomenon penalizes the brighter child by giving him a near-zero or negative change score rather than a positive one. The special experiences of the experimental children may have been sufficient to counter this trend within the group as a whole, but the trend appeared, however, to operate with the High-Criterion control group. The higher pretest scores for the SCIS criterion classrooms in both schools may reflect the fact that three out of the four had teachers who matched the curriculum's goals even before the experience started. Moreover, the sub-analysis of the children's gain differences according to socio-economic level, often taken as a rough index of intellectual capacity, showed no differences ascribable to that separation when their teacher's criterion

behavior was high. From all of these consideratins it seems likely that the differential results in this study cannot be dismissed on the basis that the more intelligent children made greater gains. In reality, intelligent children may make greater gains in the dimensions studied but the limitations of the present instrumentation and the rather homogeneous groups of subjects do not permit this explanation of the present results.

Examination of all the results show that the Internal Control measures and the shifts in categorizations of questions about the Curious Objects produced positive results for all the experimental children as compared to the control groups. Actual differences in improved cognitive fuctioning and physical science understandings were more specifically the accomplishments of the experimental, High-Criterion groups. In other words, the curriculum seemed to affect all the experimental children's attitudes toward themselves and toward adults, but the significant differences in cognitive or scientific understandings occurred only in classrooms where the teacher-child interaction was optimal in terms of the program's aims. From the point of view of developmental psychology, this finding makes good sense, since it has been repeatedly shown that "good" cognitive functioning in individuals or within situations is not possible when interaction is primarily affective. These findings suggest that the curriculum itself permitted the situation in the Low-Criterion classes to be redefined to the extent that some attitudinal shift occurred but wider gains, which included cognitive processes, occurred only in situations where the teacher, in addition to the curriculum, interpreted the situation according to SCIS aims.

DISCUSSION OF RESULTS

The failure to the Unknown Toys task to register gains seems to be due to its possible irrelevancy to children's cognitive curiosity. Examination of the absolute scores for this test in Appendix D will show that the experimental, High-Criterion class had both higher pretest and posttest scores than any other group and they they were significantly higher than their Low-Criterion counterparts. However, all classes made substantial gains so that comparisons with each other were not significant; presumably these gains were made because of the children's high interest in the "hidden" toy. Observations of the effects of this test upon the children raise an interesting question with regard to assessing gains in curiosity. If curiosity is to be measured on a pretest-posttest schedule, it may be impossible to use the same instrument both times. Familiarity reduces curiosity so that it is necessary to change the situation in some way, but the change should be minimal so that there is some certainty that the same phenomenon is being assessed. However, a minimal change is likely to cause children to hypothesize about the second situation in some quite unexpected ways. The assessment of children's questions to the Curious Objects seems to be a much better means for measuring curiosity since it evokes more thought and minimizes strong affective responses. Moreover, the content of their questions, being a richer piece of behavior than simple preference, permitted more inferences with regard to the children's hypotheses.

In summary, the results of this study suggest the following observations: (1) Some teachers match the criterion without having had the experience. The content of the SCIS criterion sort makes it clear, however, that an effective SCIS teacher is not necessarily the same as a generally effective teacher. The former is specifically involved in facilitating children's cognitive processes by confronting them with their own contradictions and disagreements, expecting and requiring that children take responsibility for their own thinking, and attempting to incite and to interest the students in cognitive, scientific pursuits; but at the same time, she expects and arranges for children to draw their own conclusions and make their own choices. The results show that some teachers move toward the SCIS criterion even during a semester's experience, but it also appears that more experience helps; some teachers with a little or even more experience do not adopt the criterion behaviors.
(2) The curriculum itself is evidently not enough to guarantee that children will gain, although the results for the Locus of Control suggest that some shift toward greater self-determination may occur in children, particularly boys, even when the experience is with a Low-Criterion teacher. Greater gains accrue for the children in the High-Criterion classrooms; they gained more in SCIS Seriation task, they conserved more, they reasoned more accurately about the effects of gravity, they came to see themselves as more self-determinative, and they became less other-dependent in their thinking and they suspected adults, correctly, of playful acts as they caught the spirit of the Curious Objects situation. A non-significant, but congruent, trend suggested that they lessened the power they ascribed to adults, if this attitude is reflected in their impression of tallness.

Although there are a number of serious limitations to this study, the convergence of results, particularly those that characterized the children in the High-Criterion experimental classrooms, and the fact that these findings were obtained by a number of different methods, designed to measure a variety of child functions (SCIS concept acquisition, cognitive processes, and attitudinal-ego functions), made it appear that the results can be taken seriously.

RECOMMENDATIONS

There are several clear recommendations that can be made: (1) Better instruments for child assessment are needed; achievement measures should have a considerable range, so that gains, unrestricted by upper limitations, can be measured; (2) More consideration needs to be given the induction and maintenance of teachers into the program; only one-half of the experimental school teachers had achieved a reasonable relationship to the criterion after their one semester's experience. The results for the children suggest that optimal gain appears when the teacher takes the criterion role; (3) Since the SCIS experience for children runs counter to the ways that many children are treated by adults, and the curriculum does not intend teachers to take a dramatic, coercive role with the children, it seems clear that its effects would best be studied on a longitudinal basis. This five-month's study showed that effects were not only specific to the curriculum but also generalizable to other important areas of children's functioning which are thought to support and permit cognitive development.

REFERENCES

Battle, E., & Rotter, J. Children's feelings of personal control as related to social class and ethnic group. Journal of Personality, 1953, 31, 482-490.

Beller, E., & Turner, J. Personality correlates of children's perception of human size. Child Development, 1964, 35, 441-449.

Block, J., & Haan, N. Ways of personality development: Continuity and change from adolescence to adulthood. New York: Appleton-Century-Crofts, in press.

Board of Education of the City of New York. Let's look at first graders. New York: BECNY, 1965.

Fahey, G. L. The questioning activity of children. Journal of Genetic Psychology, 1942, 60, 337-357.

Flanders Interaction Analysis, Paul S. Amidon & Associates, Inc., Minneapolis, Minnesota, 1963.

Haan, A., & Haan, N. A means of measuring school environments. Unpublished manuscript, San Francisco State College, 1966.

Haan, N. Curiosity. In The handbook for the assessment and measurement of educational outcomes. Berkeley: Far West Laboratory for Educational Research and Development, 1968, in preparation. Ch. 7.

Hovland, C., Lumsdaine, A., & Sheffield, F. Experiments on mass communication. Vol. 3. London: Oxford University Press, 1949. (Republished: New York, Wiley, 1965, paperback ed.)

Jersild, A. T., Thorndike, R. L., Goldman, B., & Loftus, J. J. An evaluation of aspects of the activity program in New York City public elementary schools. Journal of Experimental Education, 1939, 8, 166-207.

McNemar, Q. Psychological statistics. (3rd ed.) New York: Wiley, 1962.

Medinnus, G., & Love, J. The relation between curiosity and security in preschool children. Journal of Genetic Psychology, 1965, 107, 9-98.

Siegelman, E. Personal communication, June 1967.

Appendix A

FINAL FORM OF *Q*-SORT ITEMS

for

CLASSROOM OBSERVATION

for

SCIS STUDY

Sorting Distribution

Least Characteristic

Category	1	2	3	4	5	6	7	8	9
Number of Items	5	6	8	9	10	9	8	6	5

1. T gives detailed instructions to children on how to achieve a task.
2. T does not "fish" for pre-determined answers but is receptive to all answers based on evidence or reasoning.
3. T is irritated when a child drops, spills or knocks over things.
4. T responds to children's concerns with reassurance.
5. T expresses interest in what child was trying to do in his performance.
6. T directly sets limits for behavior during the observation period.
7. T tried to get children to memorize facts (during observation).
8. T seems readily giving, freely responding, to children who demand a lot of attention.
9. T lets children know who's boss (uses power assertion) (by threats, loud voice, etc.). (n.b. place low, if T does not employ power assertion either because she doesn't need to or because she has no authority.)
10. T takes into account children's suggestions about procedures.
11. T stays in front of classroom.
12. T reminds children of unanswered questions.
13. T shows covert hostility (e.g., coldness over meticulousness, sarcasm, etc.).
14. T uses ridicule (in form of humor) or allows pupils to ridicule mistakes of classmates.
15. T jokes and laughs with children.
16. T makes a point of summarizing what has been learned.
17. T urges children to try things on their own.
18. T is an active, lively person (e.g., teacher moves quickly, has a fast tempo). (n.b. place low if quiet, calm and slow moving.)
19. T shows approval or otherwise rewards a child's performance (n.b. place low if T more often shows disapproval.)
20. T's explanations appear to leave pupils puzzled.
21. T does not need to look at book or manual frequently.
22. T makes opportunities for children to follow their choices.

23. T seems to keep her own feelings under a tight rein. (n.b. place low if uncontrolled or expressive.)
24. T redirects a child's activity when he misbehaves.
25. T interrupts a speaking pupil.
26. T appears to have good organization in handling materials (e.g., has prepared supplies and equipment ahead of time).
27. T evokes curiosity in children beyond the intrinsic interest of material. (n.b. summary impression is called for here.)
28. T appears excited and interested in helping children make "new" discoveries, even though they are not new to her.
29. T encourages child to try again after difficulty on a task.
30. T appears to listen to children and respect their individual points of view.
31. T does not depart from usual seating arrangement. (n.b. place low if T has children grouped around her, clusters chairs for small groups, etc.)
32. T helps children to discover contradiction in their reports of observations.
33. T expresses sympathy or conveys warmth to children.
34. T makes evaluative statements comparing one child with another, or this class with another.
35. T permits children to talk among themselves as they work on these activities.
36. T permits relatively free movement among children in the classroom (e.g., children move freely about classroom as individual need arises).
37. T seems to be more comfortable with structured materials and activities than with those requiring her own or the children's improvisation. (n.b. place low if comfortable with unstructured materials.)
38. T appears certain of herself.
39. T seldom corrects child before whole class as a means of discipline.
40. T gives child reasons for doing things.
41. T appears to listen to children's responses as a guide to further discussion rather than proceeding along her own route or prescribed text.
42. T emphasizes the need for children to be orderly, neat, or quiet (e.g., silence, folding hands, sitting up straight, all eyes at me).
43. T <u>avoids</u> both positive and negative evaluations of child. (n.b. place low if uses positive and negative evaluations.)
44. T recognizes child's feelings (e.g., what he likes, dislikes, hopes and fears, needs, etc.). (n.b. place low if T gives this little attention.)
45. T responds to children's mistakes by informing then they're wrong.
46. T's face is animated when speaking to children (whether her expressiveness seems genuine or not).
47. T approaches disruptive behavior or a crisis as a problem to be worked out.(n.b. place low if simply stops, assigns, and punishes.)
48. T uses physical restraint whether punitive or not.
49. T responds to children's mistakes by asking them further questions.
50. T expresses her own feelings directly to the children by word or gesture.

51. T tried to get child to wait for what he wants.
52. T and children show physical affection for each other (e.g., touching, hugging, patting, kissing, etc.).
53. T responds to children's mistakes by giving them correct information).
54. T is irritated when a child can't find things.
55. T builds on children's disagreement to clarify conceptualizations. (n.b. place low if T ignores or stops disagreement.)
56. T and children jointly summarize their progress and/or plan their next activity. (n.b. place low if T is sole summarizer and evaluator.)
57. T responds to children's mistakes by suggesting they observe and think more.
58. T becomes rattled or angry when a child violates the rules or behaves provocatively.
59. T works with some children by themselves to individualize instruction or if child needs special help.
60. T turns children's questions to the group or the individual for further thought and observation.
61. T seems anxious to have the lesson for the day completed on time.
62. T answers simple questions of fact.
63. T is patient and continues to help even when a child does not catch on.
64. T displays a quick temper.
65. T gives detailed instructions on the mechanics of procedures.
66. T "talks down" or in a patronizing fashion to children. (n.b. place low if talks in a direct fashion to children.)

Appendix B

Reliability of Composite

for 1st and 2nd Observation*

(Spearman-Brown Corrected)

Item	1st Observation	2nd Observation
1	.20	.86
2	.25	.65
3	.41	.75
4	.77	.84
5	.28	.37
6	.47	.81
7	.39	
8	.77	.81
9	.75	.54
10	.61	.53
11	.79	.73
12		.44
13	.75	.62
14	.49	.46
15	.65	.68
16	.58	.80
17	.16	.11
18	.55	.37
19	.57	.58
20	.69	.61
21	.16	
22	.80	.73

Appendix B
(cont)

Item	1st Observation	2nd Observation
23	.46	.68
24	.69	
25	.72	
26	.73	
27	.62	
28	.69	
29	.47	
30	.76	.18
31	.28	.92
32	.41	.70
33	.77	.82
34	.72	.61
35	.77	.86
36	.41	.48
37	.73	.17
38	.68	.72
39	.64	.45
40	.53	.73
41	.87	.86
42	.48	.92
43	.89	
44	.47	.80
45	.75	.78
46	.73	.29
47	.70	.56
48	.42	.90

Appendix B
(cont)

Item	1st Observation	2nd Observation
49	.33	.53
50	.04	
51	.41	
52	.66	.71
53	.74	.78
54	.48	.81
55	.91	.61
56	.49	.61
57	.55	.11
58	.89	.28
59	.73	.65
60	.64	.41
61	.69	.87
62	.10	.72
63	.36	
64	.60	.59
65	.57	.84
66	.80	

*Items without a reliability of composite had negative correlations.

Appendix C

Results of a Tryon Cluster Analysis of Individual Observer's Sorts

	Cluster Loadings
Cluster #1 Teacher Anger	
+	
13. T shows covert hostility.	0.86
9. T lets children know who is boss (uses power assertion by threats, loud voice, etc.)	0.86
64. T displays a quick temper.	0.86
58. T becomes rattled or angry when a child violates the rules or behaves provocatively.	0.85
66. T "talks down" or in a patronizing fashion to children.	0.73
14. T uses ridicule (in form of humor) or allows pupils to ridicule mistakes of classmates.	0.70
45. T responds to children's mistakes by informing them they are wrong.	0.62
25. T interrupts speaking pupil.	0.61
34. T makes evaluative statements comparing one child with another, or this class with another.	0.60
42. T emphasizes the need for children to be orderly, neat or quiet.	0.60
48. T uses physical restraint whether punitive or not.	0.55
-	
27. T evokes curiosity in children beyond the instrinsic interest of material.	-0.75
33. T expresses sympathy or conveys warmth to children.	-0.72
28. T appears excited and interested in helping children make new discoveries, even though they are not new to her.	-0.70
30. T appears to listen to children and respect their individual points of view.	-0.70
39. T seldoms corrects child before whole class as a means of discipline.	-0.64
4. T responds to children's concerns with reassurance.	-0.61
44. T recognizes children's feelings.	-0.61
47. T approaches disruptive behavior or a crisis as a problem to be worked out.	-0.60
24. T redirects a child's activity when he misbehaves.	-0.59
22. T makes opportunities for children to follow their choices.	-0.58
43. T avoids both positive and negative evaluations of child.	-0.58
17. T urges children to try things on their own.	-0.57
19. T shows approval or otherwise rewards a child's performance.	-0.54

	Cluster Loadings
8. T seems readily giving, freely responding, to children who demand a lot of attention.	-0.50

Cluster #2 Teacher Directiveness

+

1. T gives detailed instructions to children on how to achieve a task.	0.90
65. T gives detailed instructions on the mechanics of procedures.	0.81

Cluster #3 Teacher as a Cognitive Guide

+

32. T helps children to discover contradiction in their reports of observations.	0.79
60. T turns children's questions to the group or the individual for further thought and observation	0.76
41. T appears to listen to children's responses as a guide to further discussion rather than proceeding along her own route or prescribed text.	0.76
55. T builds on children's disagreement to clarify conceptualizations.	0.59
49. T responds to children's mistakes by asking them further questions.	0.57
2. T does not "fish" for predetermined answers but is receptive to all answers based on evidence or reasoning.	0.54

-

61. T seems anxious to have the lesson for the day completed on time.	-0.58
6. T directly set limits for behavior during the observation period.	-0.52

Cluster #4 Decentralization of Teacher Effects

+

59. T works with some children by themselves to individualize instruction or if child needs special help.	0.73
35. T permits children to talk among themselves as they work on these activities.	0.59
5. T expresses interest in what child was trying to do in his performance.	0.52

	Cluster Loadings
-	
11. T stays in front of classroom.	-0.77
37. T seems to be more comfortable with structured materials and activities than with those requiring her own or the children's improvisation.	-0.59
Cluster #5 Teacher disorganized	
+	
3. T is irritated when a child drops, spills or knocks over things.	0.72
54. T is irritated when a child cannot find things.	0.67
20. T's explanation appears to leave pupil puzzled.	0.67
-	
38. T appears certain of herself.	-0.72
26. T appears to have good organization in handling materials.	0.62
Cluster #6 Teacher as an Active Person	
+	
46. T's face is animated when speaking to children whether her expressiveness seems genuine or not.	0.77
50. T expresses her own feelings directly to the children by word or gesture.	0.54
18. T is an active, lively person.	0.40
Cluster #7 Teacher as Encourager	
+	
29. T encourages child to try again after difficulty on a task.	0.56
49. T responds to children's mistakes by asking them further questions.	0.56
63. T is patient and continues to help even when a child does not catch on.	0.49
-	
53. T responds to children's mistakes by giving them correct information.	-0.61

Appendix D

Means and Standard Deviations for SCIS Tasks

a. Selection by Property

	All S's		Girls		Boys		High Criterion		Low Criterion	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Experimental										
Pretest	6.66	2.67	6.78	2.67	6.54	2.70	7.05	2.77	6.20	2.51
Posttest	8.01	2.18	8.09	2.08	7.92	2.31	8.60	2.02	7.33	2.19
Control										
Pretest	6.48	2.93	7.04	3.01	6.00	2.83	6.62	3.88	5.63	2.45
Posttest	7.93	2.66	8.56	1.99	7.41	3.06	7.92	1.68	7.52	3.19

Note: The following comparisons of mean differences were significant (significance level, *$p \leq .10$, **$p \leq .05$, ***$p \leq .01$).

Pretest:
none

Posttest:
High vs. Low Criterion (Experimental) $t = 3.10^{***}$, df,104

Appendix D
(cont)

Means and Standard Deviations for SCIS Tasks

b. Seriation

	All S's		Girls		Boys		High Criterion		Low Criterion	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Experimental										
Pretest	.04	2.05	.13	2.11	- .06	1.99	.46	2.41	- .45	1.39
Posttest	1.54	2.84	1.46	2.91	1.62	2.79	2.93	2.72	- .08	1.99
Control										
Pretest	.72	2.51	.74	2.43	.71	2.62	1.85	2.79	.52	2.53
Posttest	1.81	2.86	1.70	3.01	1.91	2.78	2.00	2.92	2.14	2.80

Note: The following comparisons of mean differences were significant (significance level, *$p \leq .10$, **$p \leq .05$, ***$p \leq .01$).

Pretest:

- Experimental vs. Control
  - (All S's) $t = -1.89^{*}$, df,162.
  - (High Criterion) $t = -1.82^{*}$, df,68.
  - (Low Criterion) $t = -2.16^{**}$, df,74.
- High vs. Low Criterion
  - (Experimental) $t = 2.32^{**}$, df,104.

Posttest:

- Experimental vs. Control
  - (Low Criterion) $t = -4.08^{***}$, df,76.
- High vs. Low Criterion
  - (Experimental) $t = 6.41^{***}$, df,104.

Appendix D
(cont)

Means and Standard Deviations for Cognitive Tasks

c. Logical Reasoning

	All S's		Girls		Boys		High Criterion		Low Criterion	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Experimental										
Pretest	22.28	4.52	22.44	4.80	22.12	4.25	24.31	3.31	19.92	4.60
Posttest	25.50	4.13	25.91	4.10	25.08	4.15	26.95	3.10	23.75	4.55
Control										
Pretest	23.68	4.75	25.44	4.06	22.24	4.85	24.62	3.12	23.45	4.65
Posttest	25.90	3.28	27.62	2.42	24.50	3.25	25.58	3.58	25.79	3.35

Note: The following comparisons of mean differences were significant (significance level, *$p \leq .10$, **$p \leq .05$, ***$p \leq .01$).

Pretest:

- Experimental vs. Control
  - (All S's) $t = -1.88^{*}$, df,164.
  - (Girls) $t = -2.79^{***}$ df,79.
  - (Low Criterion) $t = -3.26^{***}$, df,76.
- High vs. Low Criterion
  - (Experimental) $t = 5.70^{***}$, df, 104.

Posttest:

- Experimental vs. Control
  - (Girls) $t = -1.96^{*}$ df,78.
  - (Low Criterion) $t = -2.06^{**}$, df,74.
- High vs. Low Criterion
  - (Experimental) $t = 4.29^{***}$, df,104.

Appendix D
(cont)

Means and Standard Deviations for Cognitive Tasks

d. Understanding Gravity

	All S's		Girls		Boys		High Criterion		Low Criterion	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Experimental										
Pretest	3.23	1.61	3.15	1.61	3.30	1.61	3.49	1.88	2.92	1.15
Posttest	3.99	1.89	3.91	1.78	4.08	2.00	4.57	1.70	3.31	.188
Control										
Pretest	3.86	1.79	4.15	1.63	3.62	1.90	3.85	1.52	4.11	1.64
Posttest	4.55	1.85	4.33	1.57	4.74	2.07	4.00	1.73	4.78	1.99

Note: The following comparisons of mean differences were significant (significance levels, *p ≤ .10, **p ≤ .05, ***p ≤ .01).

Pretest:
- Experimental vs. Control
  - (All S's) t = -2.35**, df,163.
  - (Girls) t = -2.61**, df,78
  - (Low Criterion) t = -3.72***, df,75.
- High vs. Low Criterion
  - (Experimental) t = 1.85*, df,104.

Posttest:
- Experimental vs. Control
  - (All S's) t = -1.84*, df,163.
  - (Low Criterion) t = -3.20***, df,74.
- High vs. Low Criterion
  - (Experimental) t = 3.65***, df,105.

Appendix D
(cont)

Percentages of Children who Conserved: Cognitive Tasks

e. Conservation

	All S's		Girls		Boys		High Criterion		Low Criterion	
	+%	-%	+%	-%	+%	-%	+%	-%	+%	-%
Experimental										
Pretest	36	64	40	60	33	67	49	51	26	74
Posttest	54	46	55	45	54	46	72	28	42	58
Control										
Pretest	49	51	52	48	47	53	69	31	31	69
Posttest	54	46	56	44	53	47	46	54	50	50

Note: The following comparisons by chi-square were significant (significance levels, *$p \leq .10$. **$p \leq .05$, ***$p \leq .01$).

Pretest:
- High vs. Low Criterion
  - (Experimental) $x = 9.03$***.
  - (Control) $x = 5.21$**.

Posttest:
- Experimental vs. Control
  - (High Criterion) $x = 3.19$*.
- High vs. Low Criterion
  - (Experimental) $x = 15.64$***.

Appendix D
(cont)

Means and Standard Deviations for Locus of Control

f. Total Internal

	All S's		Girls		Boys		High Criterion		Low Criterion	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Experimental										
Pretest	4.66	1.47	4.54	1.65	4.78	1.27	4.61	1.45	4.71	1.52
Posttest	4.67	1.51	4.40	1.70	4.94	1.25	4.67	1.48	4.67	1.56
Control										
Pretest	4.69	1.54	4.44	1.65	4.90	1.42	5.92	1.26	4.19	1.57
Posttest	3.95	1.49	4.07	1.47	3.84	1.53	4.08	1.16	3.79	1.66

Note: The Following comparisons of mean differences were significant (significance levels, *p ≤ .10, **p ≤ .05, ***p ≤ .01).

Pretest:

Experimental vs. Control
(High Criterion) t =-3.01***, df,68.

High vs. Low Criterion
(Control) t = 3.48***, df,38.

Posttest:

Experimental vs. Control
(All S's) t = 2.95***, df,163.
(Boys) t = 3.62***, df,83.
(Low Criterion) t = 2.35**, df,76.

Appendix D
(cont)

Means and Standard Deviations for Locus of Control

g. Positive Internal

	All S's		Girls		Boys		High Criterion		Low Criterion	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Experimental										
Pretest	2.28	1.07	2.19	1.12	2.38	1.01	2.35	1.13	2.20	.99
Posttest	2.73	1.06	2.64	1.13	2.81	1.00	2.79	1.00	2.65	1.15
Control										
Pretest	2.56	1.11	2.56	1.22	2.58	1.03	3.08	.49	2.30	1.23
Posttest	2.32	1.18	2.70	1.17	2.00	1.11	2.58	.79	2.17	1.31

Note: The following comparisons of mean differences were significant (significance levels, *$p \leq .10$, **$p \leq .05$, ***$p \leq .01$).

Pretest:

Experimental vs. Control (High Criterion) $t = -2.27^{**}$, df,68.

High vs. Low Criterion (Control) $t = 2.19^{**}$, df,38.

Posttest:

Experimental vs. Control (All S's) $t = 2.25^{**}$, df,163.

(Boys) $t = 3.48^{***}$, df,83.

(Low Criterion) $t = 1.70^{*}$, df,76.

Appendix D
(cont)

Means and Standard Deviations for Locus of Control

h. Negative Internal

	All S's		Girls		Boys		High Criterion		Low Criterion	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Experimental										
Pretest	2.36	.99	2.33	1.13	2.39	.83	2.28	.90	2.46	1.09
Posttest	1.98	1.10	1.81	1.21	2.19	.97	1.88	1.17	2.10	1.03
Control										
Pretest	2.12	1.23	1.89	1.19	2.32	1.25	2.85	1.34	1.89	1.28
Posttest	1.64	1.13	1.37	1.28	1.87	.96	1.50	.90	1.59	1.09

Note: The following comparisons of mean differences were significant (significance levels, *p ≤ .10, *p ≤ .05, ***p ≤ .01).

Pretest:
- Experimental vs. Control
  - (High Criterion) t = -1.85*, df,68.
  - (Low Criterion) t = 2.01**, df,71.
- High vs. Low Criterion
  - (Control) t = 2.18**, df,38.

Posttest:
- Experimental vs. Control
  - (All S's) t = 1.88*, df,162.
  - (Low Criterion) t = 2.10**, df,76.

Appendix D
(cont)

Means and Standard Deviations for Other Attitudinal Tasks

i. Unknown Toys

	All S's		Girls		Boys		High Criterion		Low Criterion	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Experimental										
Pretest	4.81	3.77	4.43	3.77	5.19	3.77	6.44	3.46	2.92	3.22
Posttest	5.98	3.53	5.83	3.55	6.13	3.53	6.69	3.34	5.14	3.59
Control										
Pretest	4.27	3.52	4.15	3.48	4.38	3.61	5.23	3.61	4.39	3.59
Posttest	6.26	2.89	6.33	2.86	6.19	2.97	6.31	2.98	5.93	3.20

Note: The following comparisons of mean differences were significant (significance levels, *p ≤ .10, **p ≤ .05, ***p ≤ .01).

Pretest:

Experimental vs. Control (Low Criterion) t = -1.85*, df 75.

High vs. Low Criterion (Experimental) t = 5.39***, df,104.

Posttest:

High vs. Low Criterion (Experimental) t = 2.30**, df,105.

Appendix D
(cont)

Means and Standard Deviations for Other Attitudinal Tasks

j. Child/Adult Height Ratio

	All S's		Girls		Boys		High Criterion		Low Criterion	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Experimental										
Pretest	17.04	19.10	20.09	18.10	13.92	19.76	19.20	14.23	14.57	23.38
Posttest	17.51	24.71	19.20	20.93	15.79	28.14	23.55	18.59	10.37	29.01
Control										
Pretest	19.85	19.14	20.04	12.15	19.69	23.96	21.38	16.87	19.54	16.81
Posttest	16.32	15.85	21.92	12.23	17.47	17.17	16.17	13.48	15.85	17.05

Note: The following comparisons of mean differences were significant (significance levels, *p ≤ .10, **p ≤ .05, ***p ≤ .01).

Pretest:
none

Posttest:
High vs. Low Criterion (Experimental) t = 2.84***, df,105.

Appendix E

Comparisons of Gains between Children of White-Collar and Blue-Collar Backgrounds (Experimental High-Criterion Classes)

	White Collar $\bar{X}$	White Collar SD	Blue Collar $\bar{X}$	Blue Collar SD	t ratio	N
SCIS Tasks						
Selection by Property	2.06	3.61	1.26	3.29	.79	17 x 35
Seriation	2.24	3.05	2.68	3.44	- .45	17 x 35
General Cognitive Tasks						
Logical Reasoning	2.47	2.29	2.86	3.48	- .41	17 x 35
Understanding Gravity	.88	1.80	1.03	1.67	- .29	17 x 35
		Diff.		SE	C.R.	
Conservation		-.04		.17	-.21	
Locus of Control						
Total Internal	- .35	1.41	.12	1.43	- 1.11	17 x 35
Positive Internal	- .06	1.48	.50	1.21	- 1.44	17 x 35
Negative Internal	- .29	1.40	- .38	1.04	.25	17 x 35
Other Attitudinal Tasks						
Unknown Toys	.41	3.99	.43	3.81	- .01	17 x 35
Child/Adult Height	8.94	26.56	- .32	21.51	1.34	17 x 35

Appendix E
(cont)

Comparisons of Pretest and Post-test Scores

of Children of White-Collar and Blue-Collar Background

		White Collar $\bar{X}$	SD	Blue Collar $\bar{X}$	SD	t ratio	N
SCIS Tasks							
Selection by Property	Pre	6.71	2.73	7.31	2.75	- .75	17 x 35
	Post	8.76	2.11	8.60	2.03	.27	17 x 35
Seriation	Pre	1.12	2.69	.17	2.24	1.34	17 x 35
	Post	3.35	2.62	2.94	2.80	.51	17 x 35
General Cognitive Tasks							
Logical Reasoning	Pre	25.06	2.44	23.97	3.78	1.08	17 x 35
	Post	27.53	2.45	26.92	3.23	.69	17 x 35
Understanding Gravity	Pre	4.24	2.05	3.40	1.67	1.57	17 x 35
	Post	5.12	1.58	4.50	1.73	1.25	17 x 35
		N+	N-	N+	N-	χ^2	
Conservation	Pre	11	6	16	19	1.65	
	Post	15	2	23	12	2.95*	
Locus of Control							
Total Internal	Pre	4.41	1.06	4.83	1.64	- .95	17 x 35
	Post	4.06	1.14	4.97	1.64	- 2.06**	17 x 35
Positive Internal	Pre	2.53	1.12	2.37	1.17	.46	17 x 35
	Post	2.47	1.12	2.91	.95	- 1.49	17 x 35
Negative Internal	Pre	1.88	.86	2.46	.92	- 2.16**	17 x 35
	Post	1.59	1.28	2.06	1.16	- 1.32	17 x 35
Other Attitudinal Tasks							
Unknown Toys	Pre	6.94	2.88	6.00	3.87	.89	17 x 35
	Post	7.35	2.91	6.28	3.61	1.07	17 x 35
Child/Adult Height	Pre	19.20	20.93	21.92	12.23	.55	17 x 35
	Post	19.53	16.81	14.57	23.39	1.00	17 x 35

GPO 975-687